

UNIVERSIDADE FEDERAL DO AMAZONAS

Pró-reitora de Pesquisa e Pós-Graduação – PROPESP

Programa de Pós-Graduação em Biotecnologia - PPGBIOTEC

**Inteligência Artificial e de Dados Para Impulso À Inovação: do
Diagnóstico de Potencialidade à Transferência Tecnológica com IA
no PPGBIOTEC/UFAM**

Wilson Kume

MANAUS

2026

UNIVERSIDADE FEDERAL DO AMAZONAS

WILSON KUME

**Inteligência Artificial e de Dados Para Impulso À Inovação: do
Diagnóstico de Potencialidade à Transferência Tecnológica com IA
no PPGBIOTEC/UFAM**

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Biotecnologia (PPGBIOTEC) da Universidade Federal do Amazonas (UFAM), como parte dos requisitos necessários à obtenção do título de Doutor em Biotecnologia.

Orientadora: Prof.^a Dr.^a Maria do Perpétuo Socorro Rodrigues Chaves

Coorientador: Prof. Dr. Manoel Carlos de Oliveira Júnior

MANAUS

2026

Ficha Catalográfica

Elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

K96i Kume, Wilson
Inteligência artificial e de dados para impulso à inovação: do diagnóstico de potencialidade à transferência tecnológica com IA no PPGBIOTEC/UFAM / Wilson Kume. - 2026.
172 f. : il., color. ; 31 cm.

Orientador(a): Maria do Perpétuo Socorro Rodrigues Chaves.
Coorientador(a): Manoel Carlos de Oliveira Júnior.
Tese (doutorado) - Universidade Federal do Amazonas,
Programa de Pós-Graduação em Biotecnologia, Manaus, 2026.

1. Transferência Tecnológica. 2. Análise Envoltória de Dados (DEA). 3. Inteligência Artificial (IA). 4. Processamento de Linguagem Natural (NLP). 5. Machine Learning (ML). I. Chaves, Maria do Perpétuo Socorro Rodrigues. II. Oliveira Júnior, Manoel Carlos de. III. Universidade Federal do Amazonas. Programa de Pós-Graduação em Biotecnologia. IV. Título

UNIVERSIDADE FEDERAL DO AMAZONAS

WILSON KUME

**Inteligência Artificial e de Dados Para Impulso À Inovação: do
Diagnóstico de Potencialidade à Transferência Tecnológica com IA
no PPGBIOTEC/UFAM**

Tese para aprovação no curso de
Pós-graduação em Biotecnologia do programa
PPGBIOTEC da Universidade Federal do
Amazonas, ao qual é submetido esse trabalho.

Aprovada em: 16/03/2026.

BANCA EXAMINADORA

Prof.^a Dr.^a, Maria do Perpétuo Socorro Rodrigues Chaves

Prof.^a Dr.^a, Michele Lins Aracaty e Silva

Prof. Dr., Sandro Breval Santiago

Prof.^a Dr.^a, Ires Paula de Andrade Miranda

Prof. Dr., Dimas José Lasmar

AGRADECIMENTOS

Agradeço aos meus pais, que com amor, sacrifício e dedicação construíram as bases de tudo que sou hoje. Foram eles que me ensinaram os valores da honestidade, do respeito, do trabalho e da perseverança, guiando-me sempre pelo caminho do bem.

Muitas das dificuldades enfrentadas por eles permaneceram invisíveis aos meus olhos durante grande parte da minha vida, mas hoje reconheço que cada renúncia e cada esforço contribuíram para que eu pudesse alcançar oportunidades que culminam nesta conquista.

Este título de Doutor não representa apenas uma realização individual. É também o resultado da força, dos ensinamentos e do amor que recebi de meus pais ao longo de toda a minha trajetória.

Minha eterna gratidão.

"Knowledge is of no value unless you put it into practice."

Anton Chekhov

RESUMO

Este trabalho emprega um conjunto integrado de métodos de inteligência de dados para apoiar e promover a Transferência Tecnológica (TT) na Universidade Federal do Amazonas (UFAM), com foco no Programa de Pós-Graduação em Biotecnologia (PPGBIOTEC) e em sua inserção na bioeconomia amazônica. Utilizando o *Framework* Sequencial, um método de pesquisa estruturado como uma sequência de análises destinadas a identificar fatos e sinalizar rotas para o fomento à inovação, a primeira etapa utiliza a Análise Envoltória de Dados (DEA) para produzir um diagnóstico da eficiência relativa da UFAM, enquanto Instituição Científica, Tecnológica e de Inovação (ICT), na conversão de insumos em ativos tecnológicos. Em seguida, aplicam-se técnicas de Inteligência Artificial (IA) e Processamento de Linguagem Natural (NLP) ao conjunto de teses e dissertações do PPGBIOTEC (TEDE/UFAM), classificando-as automaticamente segundo os códigos da Classificação Internacional de Patentes (CIP) do conjunto Biotec-BR e seus respectivos Níveis de Maturidade Tecnológica (TRL). Esses procedimentos, utilizados de maneira estritamente metodológica, revelam potenciais de inovação, tendências temáticas e oportunidades de prospecção tecnológica. Os resultados revelam que 21,2% das pesquisas analisadas apresentam potencial de patenteamento, um volume expressivo que contrasta com a ausência de depósitos na base PATENTSCOPE. A predominância de estudos nos níveis TRL 1–3 (94,6%) confirma a vocação do PPGBIOTEC para a pesquisa de fronteira, enquanto os 5,4% que alcançam TRL 4+ evidenciam esforços de validação tecnológica. A análise de eficiência institucional indica que a UFAM apresenta espaço para ampliar seus resultados de *output*, tendo como *benchmarks* factíveis uma DMU Virtual composta por Instituições de Ensino Superior (IES) comparáveis. A integração dessas evidências, por meio do *Framework* Sequencial, permitiu identificar ativos estratégicos, mapear tendências temáticas e organizar informações relevantes em um modelo conceitual de *Dashboard* Interativo, que sintetiza dados de vigilância tecnológica e cooperação aberta no contexto da UFAM.

Palavras-chave: Transferência Tecnológica; Análise Envoltória de Dados (DEA); Inteligência Artificial (IA); Processamento de Linguagem Natural (NLP); *Machine Learning* (ML).

ABSTRACT

This work employs an integrated set of data intelligence methods to support and promote Technology Transfer (TT) at the Federal University of Amazonas (UFAM), focusing on the Graduate Program in Biotechnology (PPGBIOTEC) and its insertion in the Amazonian bioeconomy. Following the Sequential Framework, a research method structured as a sequence of analyses aimed at identifying facts and signaling directions for the promotion of innovation, the first stage utilizes Data Envelopment Analysis (DEA) to produce a diagnosis of the relative efficiency of UFAM, as a Scientific, Technological and Innovation Institution (ICT), in the conversion of inputs into technological assets. Then, Artificial Intelligence (AI) and Natural Language Processing (NLP) techniques are applied to the set of theses and dissertations of the PPGBIOTEC (TEDE/UFAM), automatically classifying them according to the codes of the International Patent Classification (CIP) of the Biotec-BR set and their respective Technological Maturity Levels (TRL). These procedures, employed on a strict methodological basis, uncover innovation potentials, thematic trends and opportunities for technological prospecting. The results reveal that 21.2% of the research analyzed has patenting potential, a significant volume that contrasts with the absence of deposits in the PATENTSCOPE database. The predominance of studies at TRL levels 1–3 (94.6%) confirms PPGBIOTEC's vocation for frontier research, while the 5.4% that achieve TRL 4+ show efforts at technological validation. The institutional efficiency analysis indicates that UFAM has room to expand its output results, having as feasible benchmarks a Virtual DMU composed of comparable Higher Education Institution (HEI). The integration of this evidence, through the Sequential Framework, allowed the identification of strategic assets, mapping thematic trends and organizing relevant information in a conceptual model of Interactive Dashboard, which synthesizes data from technological surveillance and open cooperation in the context of UFAM.

Keywords: Technology Transfer; Data Envelopment Analysis (DEA); Artificial Intelligence (AI); Natural Language Processing (NLP); *Machine Learning* (ML).

LISTA DE ABREVIATURAS, SIGLAS E SIGNIFICADOS

AMBDATA	Banco de Dados de Variáveis Ambientais para Modelagem de Espécies
ANVISA	Agência Nacional de Vigilância Sanitária
ARON	<i>Adaptive Resource Optimization Network</i>
BCC	Modelo Banker, Charnes e Cooper (DEA com retornos variáveis de escala)
BDTD	Biblioteca Digital Brasileira de Teses e Dissertações
BERT	<i>Bidirectional Encoder Representations from Transformers</i>
BERTimbau	Modelo de linguagem pré-treinado (<i>LLM - Large Language Model</i>)
BioBERT	<i>Bidirectional Encoder Representations from Transformers for Biomedical Text Mining</i>
Biotec-BR	CIPs de Biotecnologia
BSC	<i>Balanced Scorecard</i>
BSC-DEA	Combina o <i>Balanced Scorecard</i> (BSC) com a <i>Data Envelopment Analysis</i> (DEA)
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CCR	Modelo Charnes, Cooper e Rhodes (DEA com retornos constantes de escala)
CGen	Conselho de Gestão do Patrimônio Genético
CIP	Classificação Internacional de Patentes
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CONSUNI	Conselho Universitário (UFAM)
COVID-19	Infecção respiratória aguda, contagiosa e potencialmente grave, causada pelo vírus SARS-CoV-2
CRS	<i>Constant Returns to Scale</i> (Retornos Constantes de Escala)
CT&I	Ciência, Tecnologia e Inovação
CSV	<i>Comma-Separated Values</i>
CTA	Conhecimentos Tradicionais Associados
DEA	<i>Data Envelopment Analysis</i> (Análise Envoltória de Dados)
DEA-VRS	DEA com Retornos Variáveis de Escala
DMU	<i>Decision Making Unit</i> (Unidade de Tomada de Decisão)
DPB	Diretoria de Programas e Bolsas no País (CAPES)

EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
EPO	<i>European Patent Office</i> (Instituto Europeu de Patentes)
FAPEAM	Fundação de Amparo à Pesquisa do Estado do Amazonas
FCD	Fluxo de Caixa Descontado
FINEP	Financiadora de Estudos e Projetos
FIOCRUZ	Fundação Oswaldo Cruz
FORMICT	Formulário de Informações sobre a Política de Propriedade Intelectual das ICTs
FORTEC	Fórum Nacional de Gestores de Inovação e Transferência de Tecnologia
IA	Inteligência Artificial
IBICT	Instituto Brasileiro de Informação em Ciência e Tecnologia
ICT	Instituição Científica, Tecnológica e de Inovação
IES	Instituição de Ensino Superior
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
INPA	Instituto Nacional de Pesquisas da Amazônia
INPI	Instituto Nacional da Propriedade Industrial
IPC	<i>International Patent Classification</i> (ver CIP)
IREG	<i>IREG Observatory on Academic Ranking and Excellence</i>
LAI	Lei de Acesso à Informação (Lei nº 12.527/2011)
LATIPAT	Plataforma Latino-Americana de Patentes (Espacenet)
LDA	<i>Latent Dirichlet Allocation</i>
LLM	<i>Large Language Model</i>
LOA	Lei Orçamentária Anual
LPI	Lei da Propriedade Industrial (Lei nº 9.279/1996)
MAE	<i>Mean Absolute Error</i> (Erro Absoluto Médio)
MCASP	Manual de Contabilidade Aplicada ao Setor Público
MCTI	Ministério da Ciência, Tecnologia e Inovação
ML	<i>Machine Learning</i> (Aprendizado de Máquina)
MMA	Ministério do Meio Ambiente
NASA	<i>National Aeronautics and Space Administration</i>
NBC TSP	Normas Brasileiras de Contabilidade Aplicadas ao Setor Público
NIT	Núcleo de Inovação Tecnológica

NLP	<i>Natural Language Processing</i> (Processamento de Linguagem Natural)
OCDE	Organização para a Cooperação e Desenvolvimento Econômico
ODS	Objetivos de Desenvolvimento Sustentável (ONU)
OEPM	Oficina Española de Patentes y Marcas
O.E.	Objetivos Específicos
OI	<i>Open Innovation</i> (Inovação Aberta / Cooperação Aberta)
OMPI	Organização Mundial da Propriedade Intelectual (WIPO)
ONU	Organização das Nações Unidas
OOB	<i>Out-of-Bag</i> (erro fora da amostra)
PATENTSCOPE	Banco de dados de patentes global gratuito
PBC	<i>Public Benefit Corporation</i>
P&D	Pesquisa e Desenvolvimento
PD&I	Pesquisa, Desenvolvimento e Inovação
PDI	Plano de Desenvolvimento Institucional
PCTIS	Parque Científico e Tecnológico para Inclusão Social
PI	Propriedade Intelectual
PDF	<i>Portable Document Format</i>
PMI	Plano de Metas Institucional
PPGBIOTEC	Programa de Pós-Graduação em Biotecnologia (UFAM)
PROTEC	Pró-Reitoria de Inovação Tecnológica (UFAM)
PUBMED	Base de dados de citações e resumos de artigos científicos em biomedicina
R²	Coefficiente de Determinação
RAP-Optimizer	Otimizador de custos inteligente
RH	Recursos Humanos
RMSE	<i>Root Mean Square Error</i> (Raiz do Erro Quadrático Médio)
ROI	<i>Return on Investment</i> (Retorno sobre o Investimento)
RStudio	Ambiente de desenvolvimento integrado para a linguagem R
RUF	<i>Ranking</i> Universitário Folha
SaaS	<i>Software as a Service</i>
SCOPUS	Base de dados de resumos e citações de literatura revisada por pares
SIBBr	Sistema de Informação sobre a Biodiversidade Brasileira
STN	Secretaria do Tesouro Nacional

Super-DEA	<i>Super-efficiency Data Envelopment Analysis</i>
SVMs	<i>Support Vector Machines</i>
TEDE	Repositório de Teses e Dissertações (UFAM)
TF-IDF	<i>Term Frequency - Inverse Document Frequency</i>
THE	<i>Times Higher Education</i>
TRL	<i>Technology Readiness Level</i> (Nível de Maturidade Tecnológica)
TT	Transferência de Tecnologia
TTO	<i>Technology Transfer Office</i> (Escritório de Transferência de Tecnologia)
UESC	Universidade Estadual de Santa Cruz
UFAL	Universidade Federal de Alagoas
UFAM	Universidade Federal do Amazonas
UFBA	Universidade Federal da Bahia
UFC	Universidade Federal do Ceará
UFCG	Universidade Federal de Campina Grande
UFES	Universidade Federal do Espírito do Santo
UFG	Universidade Federal de Goiás
UFMA	Universidade Federal do Maranhão
UFMG	Universidade Federal de Minas Gerais
UFMS	Universidade Federal de Mato Grosso do Sul
UFPA	Universidade Federal do Pará
UFPB	Universidade Federal da Paraíba
UFPE	Universidade Federal de Pernambuco
UFPEL	Universidade Federal de Pelotas
UFPR	Universidade Federal do Paraná
UFRGS	Universidade Federal do Rio Grande do Sul
UFRJ	Universidade Federal do Rio de Janeiro
UFRN	Universidade Federal do Rio Grande do Norte
UFSCAR	Universidade Federal de São Carlos
UFSJ	Universidade Federal de São João del-Rei
UFSM	Universidade Federal de Santa Maria
UFU	Universidade Federal de Uberlândia
UFV	Universidade Federal de Viçosa
UnB	Universidade de Brasília

UNESP	Universidade Estadual Paulista Júlio de Mesquita Filho
UNICAMP	Universidade Estadual de Campinas
VPL	Valor Presente Líquido
VRS	<i>Variable Returns to Scale</i> (Retornos Variáveis de Escala)
VT	Vitrine Tecnológica
WIPO	<i>World Intellectual Property Organization</i>

LISTA DE QUADROS

QUADRO 1 – MATRIZ DE CONFUSÃO	38
QUADRO 2 – MÉTRICAS COMPLEMENTARES	38
QUADRO 3 – ESCALA DE LANDIS E KOCH	39
QUADRO 4 - ETAPAS DO <i>FRAMEWORK</i> SEQUENCIAL	134
QUADRO 5 - SÍNTESE DOS ACHADOS POR DIMENSÃO ESTRATÉGICA.....	143
QUADRO 6 - CONTRIBUIÇÕES.....	144

LISTA DE FIGURAS

FIGURA 1 - ESTRUTURA INTEGRADA DA PROPOSTA E FLUXO METODOLÓGICO.....	42
FIGURA 2 - DIAGRAMA DEA.....	43
FIGURA 3 - DIAGRAMA NLP.....	46
FIGURA 4 - DIAGRAMA ML.....	46
FIGURA 5 - <i>DASHBOARD</i> INTERATIVO DE COOPERAÇÃO ABERTA.....	139

SUMÁRIO

INTRODUÇÃO	20
1. REFERENCIAL TEÓRICO (CATEGORIAS PRINCIPAIS)	25
1.1. EFICIÊNCIA INSTITUCIONAL EM AMBIENTES REGIONAIS	25
1.2. MÉTRICA DA PRIMEIRA, SEGUNDA E TERCEIRA GERAÇÃO NA TRANSFERÊNCIA TECNOLÓGICA	26
1.3. A INÉRCIA INSTITUCIONAL DAS INSTITUIÇÕES CIENTÍFICAS, TECNOLÓGICAS E DE INOVAÇÃO (ICTS) PÚBLICAS E O FENÔMENO DA “ORTOTANÁSIA PATENTÁRIA”	28
1.4. COOPERAÇÃO ABERTA (OPEN INNOVATION) E A GESTÃO DO CONHECIMENTO	30
1.5. INTELIGÊNCIA DE DADOS E IA PARA A INOVAÇÃO EM BIOTECNOLOGIA	32
1.6. LACUNAS NA LITERATURA E JUSTIFICATIVA PARA O <i>FRAMEWORK</i> SEQUENCIAL PROPOSTO	34
1.7. A MATRIZ DE LAMBDA (λ), A MATRIZ DE FOLGAS (<i>SLACKS</i>) E A DMU VIRTUAL: FUNDAMENTOS E INTERPRETAÇÃO	35
1.8. COEFICIENTE KAPPA E MATRIZ DE CONFUSÃO: MÉTRICAS PARA VALIDAÇÃO DE CLASSIFICADORES	38
2. PROCEDIMENTOS METODOLÓGICOS	41
2.1. DESENHO DA PESQUISA E ABORDAGEM ÁGIL	41
2.2. <i>FRAMEWORK</i> SEQUENCIAL DE INTELIGÊNCIA DE DADOS PARA PROSPECÇÃO TECNOLÓGICA	41
2.3. DETALHAMENTO DA COLETA E DO TRATAMENTO DE DADOS	47
2.4. ESTUDOS DE CASO	48
2.5. CONSOLIDAÇÃO DAS ETAPAS METODOLÓGICAS	49
2.6. INTEGRAÇÃO E ANÁLISE DOS RESULTADOS	50
2.7. LÓCUS DA PESQUISA: UNIVERSIDADE FEDERAL DO AMAPÁ (UFAM)	51
3. RESULTADOS DA PESQUISA	54
CAPÍTULO 1 - TRANSFERÊNCIA TECNOLÓGICA: FATOS E PERCEPÇÕES SOBRE PATENTES EM BIOTECNOLOGIA NO BRASIL DE 2014 A 2023	56
1.1. INTRODUÇÃO: O CONTEXTO DA INOVAÇÃO E A NECESSIDADE DE FORTALECER A TRANSFERÊNCIA TECNOLÓGICA NA AMAZÔNIA	56
1.2. FUNDAMENTOS CONCEITUAIS: A PROPRIEDADE INTELECTUAL, OS FATOS DA BIOTECNOLOGIA E A PERCEPÇÃO DO CONHECIMENTO	57

1.3. ESTRATÉGIA DE MAPEAMENTO: FONTES, COLETA E DELIMITAÇÃO TEMPORAL PARA O ESTUDO RETROSPECTIVO (2014-2023).....	61
1.4. DINÂMICA E ANÁLISE PROSPECTIVA: O PANORAMA DE PATENTES EM BIOTECNOLOGIA E SUAS IMPLICAÇÕES PARA A GESTÃO DE INOVAÇÃO	65
1.5. CONSIDERAÇÕES FINAIS E A TRANSIÇÃO NECESSÁRIA PARA UM MODELO PRESCRITIVO	69
CAPÍTULO 2 - ARTIGO SOB TÍTULO: EFICIÊNCIA NO DESEMPENHO INOVADOR EM BIOTECNOLOGIA: UMA ANÁLISE ENVOLTÓRIA DE DADOS (DEA) NAS UNIVERSIDADES PÚBLICAS BRASILEIRAS	71
2.1. INTRODUÇÃO: EFICIÊNCIA INOVADORA E O PAPEL DA DEA NAS UNIVERSIDADES PÚBLICAS BRASILEIRAS.....	71
2.2. REFERENCIAL TEÓRICO: INOVAÇÃO, TRANSFERÊNCIA TECNOLÓGICA E DEA	73
2.3. METODOLOGIA: MODELO DEA EM DOIS ESTÁGIOS METODOLÓGICOS	78
2.4. RESULTADOS: EFICIÊNCIA INOVADORA DAS UNIVERSIDADES PÚBLICAS ANALISADAS	83
2.5. CONCLUSÃO: SÍNTESE DA EFICIÊNCIA INOVADORA NAS IES BRASILEIRAS	86
2.6. ANÁLISE COMPLEMENTAR: MATRIZ DE LAMBDA E DMU VIRTUAL DA UFAM.....	87
CAPÍTULO 3 - POTENCIAL DE PATENTEAMENTO E MATURIDADE TECNOLÓGICA: Análise Integrada com Inteligência Artificial no PPGBIOTEC/UFAM	92
3.1. INTRODUÇÃO: INTELIGÊNCIA ARTIFICIAL APLICADA À AVALIAÇÃO DE PATENTEABILIDADE E MATURIDADE TECNOLÓGICA.....	92
3.2. REFERENCIAL TEÓRICO: NLP, CIP, TRL E TRANSFERÊNCIA DE TECNOLOGIA	94
3.3. PROCEDIMENTOS METODOLÓGICOS: FLUXO DE INTELIGÊNCIA DE DADOS PARA AVALIAÇÃO DE PATENTEABILIDADE E MATURIDADE	98
3.4. RESULTADOS: DISTRIBUIÇÃO TRL, POTENCIAL DE PATENTEAMENTO E ATIVOS ESTRATÉGICOS	103
3.5. DISCUSSÃO: IMPLICAÇÕES PARA A INOVAÇÃO NA AMAZÔNIA E PARA AS DINÂMICAS INSTITUCIONAIS DA UFAM RELACIONADAS À INOVAÇÃO	107
3.6. CONSIDERAÇÕES FINAIS: SÍNTESE DO POTENCIAL DE INOVAÇÃO DO PPGBIOTEC	109

3.7. ANÁLISE COMPLEMENTAR: VALIDAÇÃO DO MODELO TRL E EVOLUÇÃO DAS MÉTRICAS DE DESEMPENHO	110
CAPÍTULO 4 - PREDIÇÃO DE ÁREAS PRIORITÁRIAS PARA BIOPROSPECÇÃO NA AMAZÔNIA: IDENTIFICAÇÃO DE <i>HOTSPOTS</i> UTILIZANDO <i>MACHINE LEARNING</i>	117
4.1. INTRODUÇÃO: O CONTEXTO DA BIOPROSPECÇÃO E A URGÊNCIA DA INOVAÇÃO PREDITIVA NA AMAZÔNIA.....	118
4.2. FUNDAMENTOS CONCEITUAIS: BIODIVERSIDADE, INTELIGÊNCIA ARTIFICIAL E MODELAGEM PREDITIVA.....	119
4.3. ESTRATÉGIA METODOLÓGICA: FONTES, COLETA E CONSTRUÇÃO DO MODELO PREDITIVO	124
4.4. DINÂMICA E ANÁLISE PROSPECTIVA: RESULTADOS DO MODELO PREDITIVO E IMPLICAÇÕES PARA A GESTÃO DA INOVAÇÃO	128
4.5. CONSIDERAÇÕES FINAIS E A TRANSIÇÃO PARA O MODELO INTEGRADOR.....	132
4.6. ANÁLISE COMPLEMENTAR: NATUREZA EXPLORATÓRIA E CRITÉRIOS DE CONFIABILIDADE LOCAL	133
CAPÍTULO 5 - <i>FRAMEWORK</i> SEQUENCIAL INTEGRADOR: <i>DASHBOARD</i> INTERATIVO E COOPERAÇÃO ABERTA	134
5.1. O <i>FRAMEWORK</i> SEQUENCIAL COMO CONTRIBUIÇÃO METODOLÓGICA.....	134
5.2. ARQUITETURA DO <i>DASHBOARD</i> INTERATIVO.....	135
5.3. FUNCIONALIDADES E NAVEGAÇÃO	136
5.4. COOPERAÇÃO ABERTA E TRANSPARÊNCIA	136
5.5. LIMITAÇÕES E PRÓXIMOS PASSOS.....	138
4. SÍNTESE E PERSPECTIVAS	140
4.1. SÍNTESE DOS ACHADOS.....	140
4.2. CONTRIBUIÇÕES DA TESE	144
4.3. PERSPECTIVAS FUTURAS	146
REFERÊNCIAS	148
ANEXO 1	167
A. RESULTADO DEA EM DUAS ETAPAS.....	167
B. MATRIZ DE LAMBDA.....	167
C. MATRIZ DE FOLGAS (<i>SLACKS</i>).....	168
ANEXO 2	169
CAPÍTULO 1 - COMPROVANTE DE PUBLICAÇÃO	169

ANEXO 3	170
CAPÍTULO 2 - COMPROVANTE DE PUBLICAÇÃO	170
ANEXO 4	171
CAPÍTULO 3 - COMPROVANTE DE PUBLICAÇÃO	171
ANEXO 5	172
CAPÍTULO 4 - COMPROVANTE DE PUBLICAÇÃO	172

INTRODUÇÃO

Nas últimas décadas, a biotecnologia assumiu uma função que transcende a mera investigação em laboratórios; a transição para economias sustentáveis, caracterizadas por baixa emissão de carbono, e a necessidade de enfrentar desafios socioambientais consolidaram essa área como pilar estratégico nos campos da saúde, da agricultura e da biorremediação. Entretanto, parcela expressiva desse potencial permanece subexplorada, em grande medida devido a restrições históricas relacionadas à infraestrutura, à logística e à gestão do conhecimento.

Na Universidade Federal do Amazonas (UFAM), essa agenda manifesta-se principalmente no Programa de Pós-Graduação em Biotecnologia (PPGBIOTEC), cuja produção científica possui relevância estratégica para a bioeconomia amazônica, enquanto a Pró-Reitoria de Inovação Tecnológica (PROTEC), na condição de Núcleo de Inovação Tecnológica (NIT) institucional, integra o contexto administrativo da política de inovação. Embora ambas as estruturas atuem de maneira complementar, a integração entre produção científica e gestão da inovação apresenta desafios que merecem investigação. Como discutido no Capítulo 1, essa lacuna impacta diretamente a capacidade institucional de transformar conhecimento em ativos tecnológicos, especialmente em áreas nas quais o caminho entre descoberta, proteção e aplicação costuma ser mais longo e complexo. A PROTEC cumpre as atribuições previstas no Marco Legal de Ciência Tecnologia e Inovação (CT&I) (BRASIL, 2018), mas sua atuação permanece concentrada na dimensão jurídico-administrativa, fenômeno recorrente entre NITs brasileiros.

Essa necessidade de integração também aparece na literatura. Moreira *et al.* (2024) destacam que universidades e NITs exercem papel decisivo ao conectar ciência e desenvolvimento socioeconômico. Almeida e Maricato (2021) reforçam que essas instituições funcionam como intermediárias essenciais entre conhecimento acadêmico e valor social ou econômico. No contexto amazônico, essa função adquire relevância adicional.

A avaliação da Transferência Tecnológica (TT) no Brasil fundamenta-se, predominantemente, em indicadores quantitativos, como número de patentes, contratos de licenciamento, e convênios. Esses indicadores são úteis, mas insuficientes para captar a complexidade dos processos inovadores. A literatura

reforça essa limitação: Holgersson e Aaboen (2019) demonstram que indicadores isolados não refletem a riqueza da TT; Lopes (2019) defende avaliações que considerem efeitos sociais, econômicos e ambientais; e Saia e Strachman (2022) destacam que, na biotecnologia, é necessário observar valor e aplicabilidade, e não apenas contagem formal. Na Amazônia, essa insuficiência é ainda mais evidente, pois métricas tradicionais não contemplam elementos como Conhecimentos Tradicionais Associados (CTA), inclusão social e fortalecimento da bioeconomia regional.

Diante desse cenário, torna-se relevante observar métodos avaliativos capazes de ir além do que já foi formalmente produzido. As chamadas Métricas de Terceira Geração são discutidas na literatura como alternativas que incorporam Inteligência de Dados e análises preditivas para avaliar maturidade tecnológica, impacto socioambiental e aderência à bioeconomia. Tais abordagens são frequentemente associadas à identificação antecipada de oportunidades de inovação, especialmente em regiões de alta complexidade territorial, como a Amazônia.

É nesse contexto que se insere a pergunta orientadora desta tese. Como a Inteligência Artificial (IA) e a inteligência de dados podem revelar, organizar e fortalecer os ativos biotecnológicos do PPGBIOTEC, ampliando sua capacidade de gerar inovação e impacto tecnológico na UFAM?

A relevância dessa pergunta é reforçada por evidências empíricas sobre as dificuldades operacionais da UFAM na gestão da inovação. Souza (2021), ao mapear fluxos administrativos da PROTEC, identifica trâmites extensos e baixa automação, aspectos que podem reduzir a velocidade de conversão de pesquisas em ativos tecnológicos. Cavalcante (2022) destaca limitações na visibilidade externa das tecnologias da UFAM e a ausência de plataformas digitais integradas. Miranda *et al.* (2024), ao comparar a estrutura da UFAM com a de outras ICTs, observam entraves burocráticos que afetam a tramitação de contratos e a negociação de parcerias. Esses estudos, em conjunto, delimitam o problema de pesquisa: a UFAM dispõe de capacidade científica, mas enfrenta gargalos institucionais que dificultam a conversão desse conhecimento em ativos tecnológicos protegidos e transferíveis.

Este estudo defende ser possível uma integração entre o PPGBIOTEC e as estruturas administrativas de inovação, por meio de análises sequenciais, que resultem em formulações capazes de fortalecer a gestão da inovação na UFAM.

Para enfrentar esse desafio, esta pesquisa adota um Framework Sequencial que integra diagnóstico de eficiência, prospecção tecnológica preditiva e consolidação dos resultados em um modelo conceitual de *Dashboard* Interativo.

O PPGBIOTEC reúne um conjunto expressivo de pesquisas registradas em teses e dissertações (TEDE/UFAM), que abrangem desde bioativos e processos industriais sustentáveis até estudos sobre ética, biodiversidade e conhecimentos tradicionais. Esse acervo constitui patrimônio intelectual estratégico para a Amazônia. No entanto, identificar quais trabalhos apresentam maior potencial de inovação ou maturidade tecnológica é um desafio complexo, especialmente diante do volume crescente de produção científica e da diversidade temática do programa. Como será aprofundado no Capítulo 3, parte desse potencial permanece pouco visível quando analisado apenas por métricas tradicionais.

A literatura recente indica que o uso de Processamento de Linguagem Natural (NLP) e o Aprendizado de Máquina (*Machine Learning* – ML) para analisar patenteabilidade e níveis de prontidão tecnológica (TRL) em trabalhos acadêmicos é um campo metodológico em expansão, com resultados promissores (Chikkamath *et al.*, 2022). Esse avanço abre espaço para abordagens ampliadas de prospecção tecnológica em ambientes de alta diversidade temática, como o PPGBIOTEC.

Esta pesquisa adota o conceito de transferência tecnológica em sentido amplo, abrangendo desde a geração do conhecimento até sua proteção, prospecção e eventual licenciamento. A Análise Envoltória de Dados (DEA) não reduz esse processo; oferece um diagnóstico inicial da eficiência relativa das instituições na geração de ativos tecnológicos. O uso de Inteligência de Dados e ferramentas de IA é estritamente metodológico visando indicar o potencial prospectivo dessas tecnologias como instrumentos de apoio à inovação (Filgueiras; Lui, 2023; Guerra, 2023).

A natureza interdisciplinar deste estudo, situada entre Gestão da Inovação, Biotecnologia e Ciência de Dados, amplia as fronteiras do conhecimento e oferece subsídios para decisões estratégicas que valorizem a biodiversidade amazônica. A tese busca disponibilizar evidências e análises que apoiem a reflexão institucional da UFAM, da PROTEC e do PPGBIOTEC acerca de suas dinâmicas e estratégias no contexto do desenvolvimento socioeconômico.

O Capítulo 1 examina o panorama nacional de patentes em biotecnologia. O Capítulo 2 aplica a DEA para avaliar a eficiência inovadora das universidades

públicas. O Capítulo 3 analisa o potencial oculto de patenteabilidade e maturidade tecnológica no portfólio do PPGBIOTEC. O Capítulo 4 explora o uso de IA para a prospecção de ativos biológicos na Amazônia. Esses resultados convergem, no Capítulo 5, para o modelo conceitual de *Dashboard* Interativo, a proposta de Cooperação Aberta e a apresentação da *Alive Thesis*.

A importância deste estudo decorre da necessidade de aprimorar a aptidão institucional da UFAM para converter o conhecimento científico em resultados tecnológicos. Essa demanda é reforçada pela realidade amazônica, marcada por limitações estruturais significativas Loureiro (2023), e dialoga com o argumento de Chang (2004), sobre a importância de desenvolver competências tecnológicas próprias. Também se alinha às diretrizes da Política de Desenvolvimento da Biotecnologia (Decreto 6.041/2007), da Lei de Inovação (10.973/2004) e do Marco Legal de CT&I (Lei 13.243/2016).

Em escala global, a metodologia proposta integra análise de eficiência (DEA) e inteligência artificial (IA/NLP), incorporando indicadores de Terceira Geração para mensurar potencial inovador, em consonância com D'Amato e Korhonen (2021) e Dasgupta (2021). No âmbito regional, a pesquisa oferece ferramentas analíticas que podem auxiliar a PROTEC na seleção de pesquisas com maior potencial de impacto, atendendo às lacunas identificadas por Zancanaro, Mathias e Hahn (2024) e dialogando com modelos de governança colaborativa (Patias; Baggio, 2022). A proposta também se harmoniza com o PDI 2016-2025 da UFAM (UFAM, 2016).

Por fim, a escolha do tema reflete a trajetória profissional do pesquisador, marcada por atuação em pesquisa, desenvolvimento, inovação e análise de dados. A motivação central reside na convicção de que a inteligência artificial pode aprimorar a gestão do conhecimento em instituições de pesquisa, ampliando seu impacto socioeconômico, perspectiva alinhada a Filgueiras e Lui (2023).

OBJETIVOS GERAIS E ESPECÍFICOS

Objetivo Geral

Analisar o potencial de promoção da transferência tecnológica na UFAM, com foco no PPGBIOTEC, empregando de Inteligência Artificial e de Dados que ofereçam subsídios para a compreensão das dinâmicas institucionais envolvidas na prospecção, avaliação e gestão de ativos tecnológicos.

Objetivos Específicos.

- I. Mapear o panorama tecnológico da biotecnologia no Brasil e no Amazonas (2014-2023), para identificar áreas de inovação, tendências setoriais e concentrações de patentes, contextualizando o posicionamento estratégico da UFAM nos cenários nacional e regional.
- II. Avaliar o posicionamento institucional da UFAM na geração de patentes, em perspectiva comparada com outras Instituições de Ensino Superior brasileiras e a Núcleos de Inovação Tecnológica do Amazonas, de modo a subsidiar ações estratégicas de prospecção e transferência de tecnologia.
- III. Identificar o potencial de inovação presente nas produções acadêmicas do PPGBIOTEC, classificando teses e dissertações segundo níveis de maturidade tecnológica e áreas de patenteabilidade, para apoiar a prospecção tecnológica institucional.
- IV. Explorar modelos preditivos para a bioprospecção de ativos biológicos na Amazônia, revelando padrões e *hotspots* de inovação que orientem a pesquisa aplicada em biotecnologia e contribuir para análises estratégicas da PROTEC.
- V. Integrar os resultados em um *Dashboard* Interativo orientado à Cooperação Aberta, reunindo dados de eficiência institucional, prospecção tecnológica e maturidade das pesquisas, de modo a apoiar a tomada de decisão da PROTEC e ampliar a transparência e a interação entre academia, governo e setor produtivo.

1. REFERENCIAL TEÓRICO (CATEGORIAS PRINCIPAIS)

Esta seção traz as categorias analíticas que fundamentam o referencial conceitual da investigação, com ênfase nos aspectos importantes para o desempenho da UFAM e do PPGBIOTEC no que tange à inovação sustentável em biotecnologia na região amazônica.

1.1. EFICIÊNCIA INSTITUCIONAL EM AMBIENTES REGIONAIS

Na Amazônia, a eficiência institucional abrange, além da produtividade, a estabilidade nas normas e o alinhamento estratégico (Kawabata; Camargo Junior, 2020; Kim; Ick, 2019). Nesse contexto, as competências desenvolvidas no PPGBIOTEC se articulam com fatores externos, entre eles o ambiente local, o suporte governamental e a infraestrutura de inovação, que operam em múltiplas camadas (AlMalki; Durugbo, 2023).

Diante da ausência de métricas absolutas de produtividade, a Análise Envolvente de Dados (DEA) tem se consolidado como ferramenta respeitável entre Instituições de Ensino Superior (IES) brasileiras, atuando como comparador e otimizador de recursos públicos (Gualandi Filho *et al.*, 2023; Melo, 2019; Pereira, 2020b). O entendimento moderno da eficiência institucional, contudo, situa a universidade no centro de um ecossistema de Inovação Aberta (*Open Innovation*). Transferir tecnologia eficazmente exige que as IES não apenas produzam conhecimento, mas também o absorvam, adaptem e compartilhem com setor privado, governo e comunidade (De Las Heras-Rosas; Herrera, 2021). A ineficiência, frequentemente detectada pela DEA, pode sinalizar dificuldades em administrar essas interações, um alerta especialmente relevante para a bioeconomia regional (Arima Júnior, 2022).

A avaliação da eficiência em programas de pós-graduação, como o PPGBIOTEC, demanda atenção especial aos resultados que geram conhecimento. Em biotecnologia, esses resultados aparecem majoritariamente como teses e dissertações, *outputs* qualitativos cujo potencial de transformação em patentes e produtos não é capturado por indicadores clássicos de eficiência (Monteiro, 2022; Vilhena, 2025). A DEA, portanto, funciona como ponto de partida para o diagnóstico, não como ponto de chegada.

Essa complementaridade entre eficiência institucional e prospecção tecnológica é o cerne do Objetivo Específico I. A prospecção setorial de patentes por meio dos códigos CIP do conjunto Biotec-BR estabelece *benchmarks* e *hotspots* tecnológicos que devem orientar a pesquisa regional (Tonolli *et al.*, 2023a; Weid *et al.*, 2018). A complexidade dos ativos amazônicos, então, requer introduzir técnicas de IA e NLP para transformar dados estatísticos em planejamento estratégico para a transferência tecnológica (Yoo; Lim; Kim, 2021b).

A DEA baseia-se nos modelos clássicos CCR (Charnes; Cooper; Rhodes, 1978) e (Banker; Charnes; Cooper, 1984), que permitem diferenciar eficiência técnica pura de eficiência de escala. Ao compreender que uma Unidade de Tomada de Decisão (DMU) pode operar próxima ao limite da ineficiência gerada por porte inadequado, a DEA fornece subsídios consistentes para estabelecer metas viáveis de aprimoramento (Silva, Alessandra Lopes da *et al.*, 2025). Por fim, comparar a eficácia geral das IES com a eficiência específica da transferência tecnológica em biotecnologia implica aprimorar práticas gerenciais e capacitar profissionais, alinhando estratégias institucionais às demandas do setor (Paletta; Chuba, 2022). A mera mensuração quantitativa, contudo, é insuficiente: a ineficiência organizacional vincula-se a fatores gerenciais como talento humano e apoio à inovação, exigindo análise multifacetada que diferencie lacunas estatísticas de lacunas de gestão (Corrêa *et al.*, 2019).

1.2. MÉTRICA DA PRIMEIRA, SEGUNDA E TERCEIRA GERAÇÃO NA TRANSFERÊNCIA TECNOLÓGICA

Métricas de Primeira Geração: a lógica da contagem

A transferência de conhecimento entre universidades e empresas exige parâmetros que conectem investimento a resultados, tais como patentes, artigos científicos e a formação de quadros qualificados (Fontes, 2020; Rodrigues *et al.*, 2020). As métricas ditas de Primeira Geração atendem a essa demanda diretamente, porém são limitadas: medem o desempenho essencialmente pela quantidade, ou seja, pelo número de patentes registradas e acordos de licenciamento (MCTI FORTEC, 2025). São úteis, mas insuficientes. Especialmente na biotecnologia amazônica, onde a megadiversidade impõe complexidades que números isolados não capturam.

Métricas de Segunda Geração: qualidade, colaboração e coinovação

A biotecnologia exige um modelo de transferência tecnológica que vá além da mera transação legal de um ativo de Propriedade Intelectual (PI). Ela depende de coinovação, conhecimento tácito e redes colaborativas. Esse contexto impulsionou o surgimento das métricas de Segunda Geração, que enfatizam a qualidade dos processos e dos relacionamentos universidade-empresa, tais como geração de receitas, criação de startups e *spin-offs* acadêmicos (Bauer *et al.*, 2020; Marchetti *et al.*, 2024). Essas métricas partem do pressuposto de que a colaboração é o motor da inovação (Guerrero; Siegel, 2024).

No entanto, mesmo essas métricas ainda são predominantemente descritivas e retrospectivas.

Métricas de Terceira Geração: impacto estratégico, valor social e bioeconomia

A avaliação em biotecnologia, considerando a grande variedade de espécies na Amazônia, não deve se limitar à lógica da rentabilidade, necessitando considerar os efeitos sociais e ambientais (Saia; Strachman, 2022). No caso do PPGBIOTEC/UFAM, essa avaliação envolve também a consideração de aspectos relacionados aos saberes tradicionais e aos princípios de desenvolvimento sustentável, inclusão social e repartição justa de benefícios previstos na legislação brasileira sobre acesso ao patrimônio genético (BRASIL, 2015).

As métricas de terceira geração viabilizam a transição de uma gestão intuitiva para decisões baseadas em dados, com a Ciência de Dados e a Inteligência Artificial atuando como instrumentos para ampliar iniciativas e identificar nichos de mercado (Raj, 2025). Essa abordagem conecta-se à Terceira Missão das universidades: integrar retornos sociais e produtivos aos processos de P&D (Compagnucci; Spigarelli, 2020). A inovação passa a ser medida pela competência de converter conhecimento científico em respostas a problemas socioeconômicos complexos, não pela quantidade de ativos gerados.

A biotecnologia envolve cadeias longas e incertas de P&D. O sucesso depende da capacidade institucional de reduzir riscos, adaptar tecnologias e priorizar ativos com potencial de mercado ou impacto social (Amry; Ahmad; Lu, 2021). O *benchmarking* e a DEA auxiliam nesse processo, orientando investimento para tecnologias alinhadas à sociobiodiversidade e à bioeconomia (Du; Seo, 2022; Tonolli *et al.*, 2023b). A UFAM, situada em ecossistema estratégico, enfrenta desafios para transformar conhecimento em bens de PI rentáveis (Paranhos *et al.*, 2021), e a qualidade da patente aparenta ser mais relevante do que a quantidade (Yoo; Lim; Kim, 2021b). A IES, ao ser tratada como DMU na DEA, tem sua eficácia avaliada na conversão de recursos humanos e financeiros em resultados de PI e TT, reforçando a necessidade de gestão orientada ao impacto (Guerrero; Urbano, 2019).

Limitações da DEA e a ponte para a Inteligência Estratégica

Uma restrição da DEA reside no fato de que esta se concentra no passado. Utiliza informações prévias e resultados que são passíveis de mensuração. Consequentemente, a DEA não é eficaz na antecipação de inovações que apresentam complexidade e múltiplos componentes. A utilização de IA e NLP contribui para uma análise mais eficaz, pela manipulação de diversos documentos, tais como dissertações, trabalhos acadêmicos e patentes. A avaliação categoriza o grau de inovação que tais documentos podem apresentar. Os critérios adotados incluem originalidade, similaridade com produtos já disponíveis no mercado e adequação às tendências globais (Yoo; Lim; Kim, 2021a).

Essa é a ponte conceitual que conecta a eficiência gerencial (DEA) à inteligência estratégica (IA/NLP), preparando o terreno para a próxima categoria do referencial teórico.

1.3. A INÉRCIA INSTITUCIONAL DAS INSTITUIÇÕES CIENTÍFICAS, TECNOLÓGICAS E DE INOVAÇÃO (ICTS) PÚBLICAS E O FENÔMENO DA “ORTOTANÁSIA PATENTÁRIA”

A inovação no setor público brasileiro depara-se com um padrão estrutural recorrente: a inércia institucional das ICTs. Apesar do aumento da produção científica

e do depósito de patentes, a transferência tecnológica segue ineficiente, não por limitação técnica, mas pela ausência de gestão estratégica da propriedade intelectual.

Estudos recentes demonstram que as ICTs acumulam patentes sem qualquer ação sistemática: sem valoração, prospecção de mercado ou articulação com empresas. Como resultado, tecnologias com potencial relevante permanecem estagnadas, sem retorno à sociedade que as financiou.

Menezes *et al.* (2025) apontam que a maioria das ICTs públicas não firma contratos de transferência, mesmo dispondo de ativos protegidos. Santos *et al.* (2024) acrescentam que falta avaliação econômico-financeira para orientar decisões, sem valoração, não há negociação; sem negociação, não há transferência. Complementarmente, Menezes *et al.* (2026) destacam a falta de transparência e divulgação padronizada, afastando potenciais parceiros, enquanto Lessa, De Souza e Silva (2024) mostram que a maioria das ICTs sequer reconhece esses ativos em seus balanços, reforçando a máxima de que não se pode gerir o que não se conhece.

Embora a literatura utilize termos como “patentes zumbi” ou “patentes de prateleira” para descrever esse cenário, este trabalho propõe o conceito de Ortotanásia Patentária. O termo deriva de “ortotanásia”, do grego *orthós* (correto) e *thanatos* (morte), empregado na bioética para designar a morte natural, sem prolongamento artificial. Por analogia morfológica e conceitual, Ortotanásia Patentária refere-se ao fenômeno em que a instituição detentora de um ativo tecnológico mantém sua vigência jurídica por meio do pagamento de anuidades, mas não desenvolve esforços sistemáticos para sua transferência. Nessa condição, o ativo permanece custodiado até o decurso do prazo legal que leva à sua extinção (caducidade).

O conceito aqui introduzido oferece uma lente analítica para compreender a desconexão entre o esforço de proteção intelectual e a efetiva entrega de inovação. A Ortotanásia Patentária é, portanto, o sintoma de uma estrutura de gestão reativa e burocrática. O prejuízo é concreto: tecnologias capazes de gerar emprego e competitividade permanecem circunscritas ao ambiente acadêmico, perecendo lentamente por inanição comercial.

Além da ortotanásia patentária, a literatura aponta que a atuação dos NITs brasileiros tende a concentrar-se na dimensão jurídico-administrativa, com baixa ativação tecnológica. Almeida e Maricato (2021) afirmam que, embora os NITs sejam concebidos como formas de organização estratégica, sua atuação se dá, predominantemente, por meio de práticas burocráticas. Conforme a pesquisa

elaborada por Moreira *et al.* (2024), a conversão do conhecimento em valor tecnológico acontece, predominantemente, em razão da carência de métodos que se mostrem efetivamente eficientes na promoção de incentivos. Este estudo apresenta uma interpretação suplementar para tal fenômeno ao examinar o arcabouço legal de Ciência, Tecnologia e Inovação como um conjunto de *lex imperfecta*. Normas como a Lei de Inovação (10.973/2004), o Marco Legal de Ciência, Tecnologia e Inovação (13.243/2016) e a Política de Desenvolvimento da Biotecnologia (Decreto 6.041/2007) definem responsabilidades institucionais, entretanto, tais diretrizes não abordam sanções para aqueles que não as observarem. Dessa forma, contribui-se para que as instituições dedicadas à ciência e à tecnologia permaneçam inativas, dedicando-se mais a cumprir as formalidades do que a converter o conhecimento em tecnologias aplicáveis. Os dados nacionais do Índice Brasil de Inovação e Desenvolvimento (INPI, 2025) sustentam essa leitura ao evidenciar assimetrias estruturais entre insumos e resultados, isto é, a fragilidade institucional e a baixa conversão tecnológica são fenômenos sistêmicos no país.

Essa constatação suscita a necessidade premente de instrumentos que permitam diagnosticar a eficiência institucional e identificar potenciais de inovação ainda latentes. É nesse ponto que se justificam as abordagens adotadas nesta tese. A DEA, as técnicas de NLP, a classificação por TRL e os modelos de Inteligência Artificial emergem como ferramentas para superar os gargalos da Ortotanásia Patentária e apoiar as estruturas administrativas de inovação da UFAM, na gestão estratégica da inovação.

1.4. COOPERAÇÃO ABERTA (OPEN INNOVATION) E A GESTÃO DO CONHECIMENTO

A Cooperação Aberta (*Open Innovation* - OI) ultrapassa o modelo convencional de inovação restrita, ao criar fluxos bidirecionais de saber entre universidades, empresas, governo e sociedade (Kissimoto; Mattos; Laurindo, 2023). Para as IES, a implementação referida atende à exigência por uma maior relevância socioeconômica e à otimização da transferência de tecnologias, especialmente em cenários de recursos limitados, nos quais a inovação deve ajustar-se às particularidades regionais (Leão *et al.*, 2024).

O conceito alinha-se ao modelo da Hélice Quíntupla (Quintuple Helix), que amplia a tríplice hélice (academia-indústria-governo) para incluir a sociedade civil e o meio ambiente como esferas estruturantes dos sistemas de inovação (Carayannis; Barth; Campbell, 2012). Na Amazônia, essa abordagem revela-se vital: permite que comunidades tradicionais, detentoras de conhecimentos associados, participem do desenvolvimento de soluções alinhadas à repartição de benefícios (Saia; Strachman, 2022).

A OI afeta diretamente a eficiência mensurada pela DEA. Colaborações geram *inputs* não financeiros, acesso à infraestrutura, *expertise* empresarial, inteligência de mercado, e qualificam os *outputs*, elevando a qualidade da transferência tecnológica e o impacto socioeconômico (Fáveri, 2020). No entanto, a efetividade da OI esbarra na capacidade institucional de gerir volume e heterogeneidade informacionais. A inteligência de dados torna-se, então, diferencial estratégico.

Tecnologias colaborativas, como *dashboards* interativos, convertem indicadores de desempenho e parcerias em representações visuais acessíveis, transformando a OI de intenção em processo gerencial guiado por dados (Kissimoto; Mattos; Laurindo, 2023). A capacidade de colaborar configura-se, assim, como um indicador de desempenho gerencial: DMU engajada em arranjos de OI tende a ser mais adaptável e estrategicamente alinhada.

Para que a OI no PPGBIOTEC seja bem-sucedida, a UFAM necessita de capacidade absorptiva, ou seja, a habilidade de identificar, absorver e utilizar conhecimento externo, bem como de capacidade disseminativa, eficácia em transformar conhecimento interno em resultados transferíveis (Tonial *et al.*, 2022). Na biotecnologia, os desafios são amplificados pelos longos ciclos de maturação e pelo intenso escrutínio regulatório de órgãos como Agência Nacional de Vigilância Sanitária (ANVISA) ou do Conselho de Gestão do Patrimônio Genético (CGen), exigindo gestão de conhecimento integrada e seletiva (Bashir, 2024).

Plataformas digitais colaborativas configuram um caminho possível para codificar o conhecimento tácito dos pesquisadores em ativos explícitos, viabilizando sua análise por inteligência de dados (Paiva; Ribeiro; Coutinho, 2020). Sob uma ótica de análise de Terceira Geração, a rede colaborativa que a OI tece constitui, por si só, um *output* estratégico: a posição da universidade em redes de inovação prediz sucesso futuro em transferência de tecnologia com mais eficácia do que métricas tradicionais, como número de patentes.

A sinergia entre Cooperação Aberta e bioeconomia é notável: ambas se concentram na geração de valor sustentável a partir de recursos biológicos. A governança colaborativa evita as chamadas "inovações de gaveta" e assegura que a eficiência do PPGBIOTEC esteja alinhada aos Objetivos de Desenvolvimento Sustentável (ODS) (Bartz *et al.*, 2020).

Por fim, para incorporar a OI à avaliação DEA, *inputs* colaborativos podem ser tratados como variáveis contextuais não discricionárias. Modelos como o BSC-DEA¹ permitem separar a eficiência interna de gestão (controlável pelo NIT) da eficiência de engajamento (relacionada ao ambiente de OI), viabilizando um diagnóstico mais preciso do desempenho do PPGBIOTEC no ecossistema amazônico (Najafi *et al.*, 2023).

1.5. INTELIGÊNCIA DE DADOS E IA PARA A INOVAÇÃO EM BIOTECNOLOGIA

Segundo o Manual de Oslo, a inovação configura-se como a implementação de um produto, processo ou modelo novo, ou que tenha sido significativamente aperfeiçoado (OCDE, 2018). Na biotecnologia, emprega-se o conhecimento sobre biologia e biomoléculas para criar produtos de valor, principalmente aqueles provenientes de recursos genéticos e bioativos (OCDE, 2009).

A principal restrição da DEA reside na sua fundamentação em resultados previamente obtidos, como as patentes concedidas. A administração da inovação na biotecnologia deve reconhecer o potencial inovador antes de sua materialização. A Inteligência de Dados e a Inteligência Artificial tornam-se fundamentais nesse contexto.

No campo da propriedade intelectual, a Inteligência de Dados extrapola a simples contagem de documentos. Inclui-se o exame de teses, dissertações, artigos e patentes. Dessa forma, a inteligência de dados, converte extensas quantidades de dados em conhecimento capaz de subsidiar projeções futuras. A Inteligência Artificial, em particular o Processamento de Linguagem Natural, possibilita extrair informações de textos extensos. Yoo, Lim e Kim (2021a) demonstraram que os modelos de aprendizado profundo realizam a análise de patentes a fim de avaliar sua qualidade e

¹ BSC-DEA integração do Balanced Scorecard (BSC) com a Data Envelopment Analysis (DEA) para uma avaliação de desempenho mais robusta.

capacidade inventiva. Esse é um aspecto que não é administrado pela DEA. É necessário compreender que as maneiras de operar com a linguagem demandam o uso de termos complexos. Isso dificulta a compreensão dos resultados apresentados.

A biotecnologia contribui para a aplicação da inteligência artificial. Isto ocorre em razão da combinação de dados estruturados e não estruturados. A inteligência artificial identifica padrões de pesquisa. Ela examina as semelhanças entre distintos projetos. Ademais, indica as lacunas presentes no mercado em termos de opções (Mishra; Tiwari, 2024). Este trabalho apresenta um modelo sequencial que combina a DEA com IA. A DEA analisa a eficácia do NIT. A inteligência artificial avalia se o material do Programa de Pós-Graduação em Biotecnologia é inovador. A elevada aptidão para gerar inovações, aliada à baixa eficiência na DEA, indica a existência de um problema relacionado à administração.

A precisão dos modelos de processamento de linguagem natural está condicionada à qualidade satisfatória do *corpus* textual empregado para o aprendizado. A metodologia de trabalho necessita descrever minuciosamente o processo de coleta e a transformação dos textos. A Inteligência de Dados proporciona maior transparência e organização no planejamento. Ela auxilia as Instituições de Ensino Superior na transformação de dados científicos em informações que contribuem para a gestão (Kissimoto; Mattos; Laurindo, 2023). Na Amazônia, a inteligência artificial pode auxiliar na seleção e na síntese de substâncias com maior celeridade. Isso resulta em economia de tempo e recursos financeiros (Holzinger *et al.*, 2023).

A análise sobre a propriedade intelectual por meio de NLP identifica indícios de inovação em locais pouco usuais, como sínteses de projetos. O NIT pode avaliar o montante de um ativo antes de atingir sua maturidade. Esse fenômeno ocorre ao converter saberes não documentados em dados de pesquisa claramente (Hassani *et al.*, 2020). O *Framework* Sequencial proposto, ao combinar DEA e IA, expõe gargalos da cadeia de valor da inovação, como a morosidade na elaboração de patentes e as dificuldades de negociação de licenças (Zhang *et al.*, 2024).

A Inteligência de Dados adota o modelo da Hélice Quíntupla. Esse modelo contribui para introduzir inovações na sociedade. Dessa forma, a DEA passa a constituir um instrumento de responsabilidade da Terceira Geração (Carayannis; Preissler, 2025).

Igualmente relevante é a atenção às questões éticas relacionadas à privacidade dos indivíduos e à gestão do meio ambiente (Omrow *et al.*, 2024).

1.6. LACUNAS NA LITERATURA E JUSTIFICATIVA PARA O *FRAMEWORK* SEQUENCIAL PROPOSTO

A avaliação de eficiência de ICTs e a prospecção tecnológica fundamentada em inteligência artificial são temas recorrentes, embora tratados fragmentadamente, na literatura. A DEA, aplicada à eficiência institucional no trabalho de Pereira (2020b), contrasta com o uso de técnicas de NLP para a análise de patentes, como demonstrado por Yoo, Lim e Kim (2021a). No entanto, não foram identificadas propostas que integrem esses métodos em um fluxo sequencial capaz de conectar diagnóstico macro de eficiência à prospecção micro de ativos tecnológicos em textos acadêmicos e, subsequentemente, à predição de *hotspots* de inovação.

Essa ausência revela uma lacuna metodológica relevante: embora existam ferramentas isoladas para medir desempenho ou extrair conhecimento de bases textuais, não há modelos que articulem essas etapas maneira sistemática, permitindo que a DEA funcione como filtro e insumo para etapas posteriores de NLP e ML. A literatura tende a ser descritiva, focada em medir o que já ocorreu, mas raramente oferece mecanismos prescritivos que orientem o que pode ser transferido ou desenvolvido.

Além disso, os estudos tradicionais de inovação e transferência tecnológica permanecem ancorados em métricas de primeira e segunda geração. O *framework* proposto nesta tese aponta na direção de métricas de terceira geração, centradas na identificação de “ativos invisíveis”, conhecimentos latentes em teses, dissertações e artigos que ainda não foram convertidos em propriedade intelectual formal, mas apresentam alto potencial de mercado.

Outro aspecto ausente na literatura é a integração entre métodos computacionais e mecanismos de redução da assimetria de informação entre ICTs e setor produtivo. A proposta desta tese incorpora um *Dashboard* Interativo e um modelo de Cooperação Aberta como etapa final do *framework*, permitindo que os resultados de IA sejam traduzidos em indicadores de inovação e transferência tecnológica.

Por fim, a validação do modelo no contexto da bioeconomia amazônica, especificamente no PPGBIOTEC/UFAM, indica que o *framework* não apenas é conceitualmente inovador, mas também aplicável e estratégico, permitindo identificar ativos biotecnológicos com maior celeridade e orientada por dados do que métodos empíricos tradicionais.

Dessa forma, o presente trabalho preenche uma lacuna teórica e metodológica, oferecendo às ICTs um roteiro replicável, baseado em plataforma de *open science* e bases públicas, que supera a gestão reativa da propriedade intelectual e inaugura uma abordagem de prospecção ativa orientada por inteligência de dados.

A operacionalização desse *framework* é detalhada na seção metodológica (Capítulo 3), sendo aplicada nos Capítulos 1 a 4 e integrada no *Dashboard* Interativo (Capítulo 5).

1.7. A MATRIZ DE LAMBDA (λ), A MATRIZ DE FOLGAS (*SLACKS*) E A DMU VIRTUAL: FUNDAMENTOS E INTERPRETAÇÃO

A DEA gera um conjunto de informações que explora o escore de eficiência tradicionalmente utilizado em comparações entre instituições. Entre esses resultados complementares, destacam-se a matriz de lambdas (λ), a matriz de folgas (*slacks*) e a DMU Virtual, elementos que permitem compreender não apenas o quanto uma unidade se encontra distante da fronteira eficiente, mas também de que maneira essa distância se manifesta e quais caminhos são factíveis para sua superação. Cooper, Seiford e Tone (2007) observam que esses componentes revelam a estrutura interna da eficiência, oferecendo uma leitura mais rica do que aquela obtida apenas pelo valor escalar do índice. Embora não ampliem a DEA do ponto de vista metodológico, fortalecem sua utilização estratégica, especialmente quando se busca orientar políticas de inovação e apoiar decisões institucionais.

A formulação essencial da DEA orientada a *output*, sob retornos variáveis de escala (VRS), pode ser expressa da seguinte forma para uma DMU:

$$\max_{\theta, \lambda, s^+, s^-} \theta$$

sujeito a:

$$\begin{aligned}
\sum_{j=1}^n \lambda_j x_{ij} &\leq x_{i0} - s_i^- \forall i \\
\sum_{j=1}^n \lambda_j y_{rj} &\geq \theta y_{r0} + s_r^+ \forall r \\
\sum_{j=1}^n \lambda_j &= 1 \\
\lambda_j &\geq 0, s_i^- \geq 0, s_r^+ \geq 0
\end{aligned}$$

Essa formulação é bastante adequada para embasar a interpretação dos resultados apresentados nesta tese, dispensando um aprofundamento em derivadas matemáticas que fogem ao escopo do trabalho.

Os lambdas (λ) representam os pesos atribuídos às DMUs eficientes que compõem a fronteira de referência para a unidade avaliada. Charnes, Cooper e Rhodes (1978) destacam que esses coeficientes expressam a “composição” da fronteira eficiente, funcionando como um mecanismo de comparação estrutural. Na prática, os λ mostram quais instituições funcionam como *benchmarks* realistas e a proporção de contribuição de cada uma para a referência. Isso revela semelhanças institucionais e permite encontrar modelos de desempenho que podem ser alcançados no grupo analisado.

A matriz de folgas complementa essa interpretação ao indicar, diretamente, quais variáveis precisam ser ajustadas e em que magnitude para que a DMU alcance sua referência. As folgas são definidas por:

$$\begin{aligned}
s_i^- &= x_{i0} - \sum_{j=1}^n \lambda_j x_{ij} \\
s_r^+ &= \sum_{j=1}^n \lambda_j y_{rj} - \theta y_{r0}
\end{aligned}$$

Thanassoulis, Portela e Despić (2008) enfatizam que as folgas são essenciais para distinguir entre ineficiências atribuíveis à escala, à má alocação de insumos ou à insuficiência de *outputs*. Elas permitem identificar se uma unidade apresenta excesso

relativo de insumos, déficit específico de resultados ou ambos, oferecendo uma visão detalhada da natureza da ineficiência.

A DMU Virtual, por sua vez, é a combinação ponderada das DMUs eficientes que formam a referência para a unidade avaliada. Ela pode ser expressa como:

$$x_i^{virtual} = \sum_{j=1}^n \lambda_j x_{ij}$$

$$y_r^{virtual} = \sum_{j=1}^n \lambda_j y_{rj}$$

Cook e Zhu (2008) descrevem a DMU Virtual como uma “imagem espelhada” da fronteira eficiente, representando um padrão de desempenho simultaneamente ideal e factível. Ela não constitui um modelo hipotético, mas uma síntese concreta do que a instituição poderia alcançar caso se aproximasse de seus *benchmarks* naturais.

A integração desses elementos transforma a DEA de um diagnóstico estático, frequentemente interpretado como uma classificação entre eficientes e ineficientes, em um instrumento prescritivo. A matriz de lambdas, a matriz de folgas e a DMU Virtual, quando interpretadas conjuntamente, permitem identificar trajetórias de melhoria, reduzir interpretações equivocadas sobre rankings e oferecer insumos para a gestão estratégica, especialmente para núcleos de inovação tecnológica que precisam compreender não apenas o desempenho atual, mas também as possibilidades de evolução institucional. Dyson *et al.* (2001) argumentam que a DEA atinge seu potencial máximo quando seus resultados são analisados integradamente, e não isolada.

Apesar de sua utilidade, esses resultados devem ser interpretados com cautela. Simar e Wilson (2007) destacam não haver inferência causal, que múltiplas soluções ótimas podem existir e que a seleção de inputs e *outputs* influencia fortemente os resultados. Além disso, aspectos qualitativos da inovação, como impacto social, maturidade tecnológica ou aderência à bioeconomia, não são capturados pela DEA, exigindo métodos complementares. Por essa razão, nesta tese, a matriz de lambdas, a matriz de folgas e a DMU Virtual são utilizadas como elementos complementares, e não como núcleo interpretativo. Elas funcionam como ponte entre o diagnóstico inicial e as etapas subsequentes do *Framework* Sequencial, que

incorporam inteligência artificial, análise de maturidade tecnológica e integração em um *dashboard* interativo.

1.8. COEFICIENTE KAPPA E MATRIZ DE CONFUSÃO: MÉTRICAS PARA VALIDAÇÃO DE CLASSIFICADORES

A avaliação de desempenho de modelos de classificação supervisionada, como o *Random Forest* utilizado no Capítulo 3, exige métricas que vão além da simples acurácia. Duas ferramentas analíticas se destacam nesse contexto: a matriz de confusão e o coeficiente Kappa de Cohen.

Matriz de Confusão

A matriz de confusão é uma tabela que confronta as classes preditas pelo modelo com as classes reais (observadas). Sua estrutura fundamental é:

Quadro 1 – Matriz de Confusão

n x n	Predito Classe A	Predito Classe B
Real Classe A	Verdadeiro Positivo (VP)	Falso Negativo (FN)
Real Classe B	Falso Positivo (FP)	Verdadeiro Negativo (VN)

Fonte: Elaborado pelo autor (2026).

A partir dessa matriz, derivam-se métricas complementares:

Quadro 2 – Métricas Complementares

Métrica	Cálculo	Descrição
Acurácia	$(VP + VN) / (VP + FN + FP + VN)$	proporção de acertos totais
Sensibilidade	$VP / (VP + FN)$	capacidade de identificar positivos
Especificidade	$VN / (VN + FP)$	capacidade de identificar negativos
Precisão*	$VP / (VP + FP)$	exatidão das predições positivas.

Fonte: Elaborado pelo autor (2026).

Para problemas multiclasse, como a classificação TRL, a matriz de confusão expande-se para $k \times k$ dimensões, onde k é o número de classes (Congalton, 1991; Stehman, 1997).

Coeficiente Kappa de Cohen

O coeficiente Kappa (κ) é uma medida de concordância entre duas avaliações que corrige o acordo esperado pelo acaso. Proposto por Cohen (1960), é amplamente utilizado em estudos de validação de classificadores, especialmente quando há desbalanceamento entre classes.

A fórmula do Kappa é:

$$\kappa = \frac{p_0 - p_e}{1 - p_e}$$

p_0 — Concordância Observada é a proporção de casos em que ambos os avaliadores concordaram efetivamente (acurácia).

p_e — Concordância Esperada é a proporção de concordância esperada pelo acaso, calculada como o produto das probabilidades marginais.

(Landis e Koch (1977) propuseram uma escala de interpretação amplamente adotada:

Quadro 3 – Escala de Landis e Koch

Kappa (κ)	Interpretação
< 0,00	Pior que o acaso
0,00 - 0,20	Concordância leve (<i>slight</i>)
0,21 - 0,40	Concordância razoável (<i>fair</i>)
0,41 - 0,60	Concordância moderada (<i>moderate</i>)
0,61 - 0,80	Concordância substancial (<i>substantial</i>)
0,81 - 1,00	Concordância quase perfeita (<i>almost perfect</i>)

Fonte: Elaborado pelo autor (2026).

Aplicação na Validação de Classificadores TRL

Na avaliação de modelos de classificação de Níveis de Maturidade Tecnológica (TRL) em textos acadêmicos, a matriz de confusão e o Kappa são particularmente relevantes por três razões:

- I. Classes desbalanceadas: a distribuição natural dos TRLs em programas de pós-graduação não é uniforme;

- II. Similaridade linguística: resumos de diferentes TRLs compartilham vocabulário técnico semelhante;
- III. Custo do erro: aqueles entre classes adjacentes (ex: TRL 3 ↔ TRL 4) são menos graves que erros entre classes distantes (ex: TRL 1 ↔ TRL 5).

A literatura recomenda o uso combinado da matriz de confusão (para análise qualitativa dos erros) e do Kappa (para avaliação quantitativa da concordância) como práticas padrão em estudos de classificação textual (Congalton, 1991; Foody, 2002).

Limitações e Cuidados Metodológicos

O Kappa apresenta algumas limitações que devem ser consideradas:

- Prevalência: quando uma classe é muito rara ou muito frequente, o Kappa pode ser paradoxalmente baixo mesmo com alta acurácia;
- Dependência do número de classes: o valor máximo do Kappa diminui à medida que o número de classes aumenta;
- Interpretação contextual: a escala de Landis e Koch (1977) é uma referência, não um critério absoluto.

Por essas razões, recomenda-se que o Kappa seja sempre apresentado em conjunto com a matriz de confusão, permitindo uma compreensão mais rica do desempenho do classificador (Banerjee *et al.*, 1999; Viera; Garrett, 2005).

2. PROCEDIMENTOS METODOLÓGICOS

2.1. DESENHO DA PESQUISA E ABORDAGEM ÁGIL

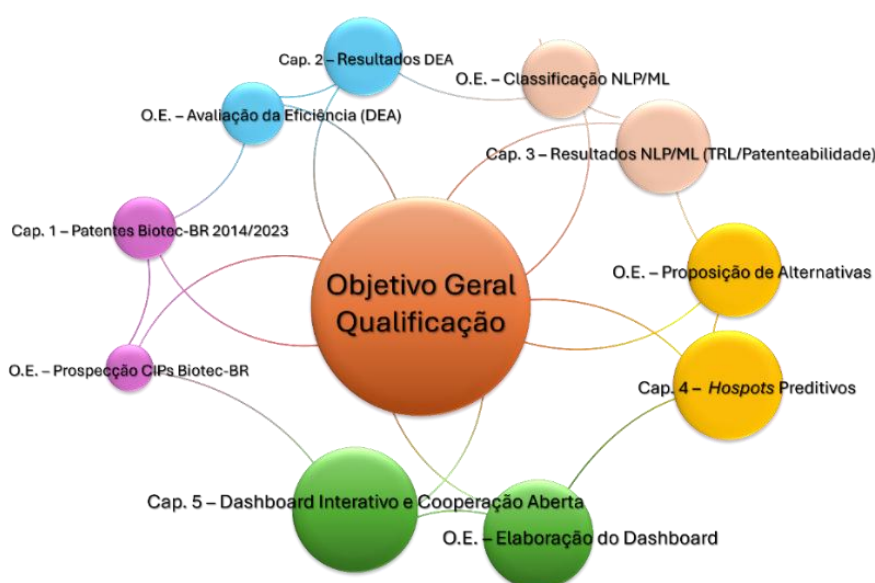
A investigação, na fase quantitativa, mapeia a posição da UFAM no espectro da eficiência por meio da DEA; na fase qualitativa, emprega IA/NLP, em uma arquitetura sequencial explanatória. A etapa qualitativa investiga por que não se está mais longe e onde está o potencial inovador do conhecimento. Trata-se de um estudo minucioso e de caso sobre o Programa de Pós-Graduação em Biotecnologia da UFAM. Ela compara seu desempenho com *benchmarks* de outros PPGs de Biotecnologia localizados na Região Norte e nas principais IES do Brasil. A pesquisa se beneficia da adoção dos princípios de metodologias ágeis, permitindo ajustes contínuos e colaboração entre áreas ao longo do desenvolvimento. Hidalgo (2018) aponta que os métodos ágeis são ideais para gerenciar projetos que envolvem diversas disciplinas e para promover mudanças em áreas de inovação tecnológica, como a bioeconomia.

2.2. *FRAMEWORK* SEQUENCIAL DE INTELIGÊNCIA DE DADOS PARA PROSPECÇÃO TECNOLÓGICA

Este estudo utiliza um *Framework* Sequencial, organizado em cinco etapas interligadas, nas quais os resultados obtidos em cada fase são utilizados na próxima, constituindo um processo contínuo de diagnóstico e prospecção tecnológica. As etapas são: (i) levantamento do panorama tecnológico de patentes em biotecnologia no Brasil; (ii) avaliação da eficiência institucional na conversão de recursos em inovação; (iii) identificação automatizada do potencial de patenteamento e da maturidade tecnológica (TRL) em teses e dissertações; (iv) predição de áreas prioritárias para bioprospecção na Amazônia; e (v) consolidação dos resultados em um modelo conceitual de *Dashboard* Interativo orientado à vigilância tecnológica e à cooperação aberta. A Figura 1 ilustra as interconexões entre cada Objetivo Específico (O.E.) e o respectivo Capítulo (Cap.), enfatizando a coesão do percurso analítico e evidenciando como o *Framework* estrutura e integra o desenvolvimento da tese.

O *Framework* Sequencial não se configura como um produto ou ferramenta, mas como um método de construção da tese: uma série de ações e investigações que se destinam a identificar fatos e a sinalizar caminhos para impulsionar a inovação.

Figura 1 - Estrutura Integrada da Proposta e Fluxo Metodológico.



Fonte: Elaborado pelo autor (2025).

Análise de Envoltória de Dados (DEA): para a avaliação da eficiência

Os *inputs* e *outputs* utilizados no modelo DEA já foram definidos e justificados no Capítulo 2 (Seção 2.3 - METODOLOGIA: MODELO DEA EM DOIS ESTÁGIOS METODOLÓGICOS), onde se encontram detalhados os insumos institucionais considerados (*inputs*) e os resultados inovadores observáveis (*outputs*). No presente capítulo, a DEA é retomada apenas como etapa integrante do *Framework* Sequencial, utilizando os resultados previamente obtidos para compor o diagnóstico inicial de eficiência institucional, conforme previsto nos objetivos específicos destacados na Seção 1.

A DEA avalia a eficiência relativa de unidades comparáveis, identificando *benchmarks* e apontando onde há espaço para melhoria.

Essa fronteira age como um envelope que contém o melhor desempenho já registrado no grupo em questão. As DMUs que se encontram sobre essa fronteira são vistas como eficientes (*benchmarks*), enquanto as que estão abaixo dela têm espaço

para melhorias e, portanto, são classificadas como ineficientes em relação às suas pares. Essa comparação é multidimensional, ou seja, é possível comparar vários insumos e produtos ao mesmo tempo, mesmo que sejam de naturezas diferentes ou de escalas distintas.

No contexto desta pesquisa, a DEA será aplicada em sua formulação orientada a *outputs*, buscando identificar quais instituições conseguem maximizar resultados em inovação, incluindo patentes e outros indicadores de desempenho inovador definidos no Capítulo 2, a partir de um conjunto dado de insumos. Essa formulação orientada a *outputs* é adequada quando o interesse institucional está em ampliar os resultados obtidos sem a necessidade de expandir proporcionalmente os insumos disponíveis.

Figura 2 - Diagrama DEA



Fonte: Elaborado pelo autor (2025).

Papel da DEA no *Framework* Sequencial

No contexto do *Framework* Sequencial, a DEA cumpre a função de etapa diagnóstica. A partir dos *inputs* e *outputs* definidos no Capítulo 2, o modelo calcula a eficiência relativa das IES comparadas, permitindo identificar gargalos institucionais e assim atender ao Objetivo Específico 1 (avaliar a eficiência inovadora da UFAM). Os resultados da DEA servem como base para a próxima etapa (IA/NLP), que expande a análise qualitativa e preditiva do potencial de inovação do PPGBIOTEC.

Limitações Críticas da DEA.

A DEA apresenta limitações conhecidas, como sensibilidade a dados históricos e à seleção de variáveis, conforme discutido no Capítulo 2.

A literatura menciona essas fragilidades metodológicas, destacando a sensibilidade da DEA e dos modelos de IA a dados desbalanceados e a ruídos estatísticos (Stakoviak, 2023).

Estratégias de Mitigação de Vieses.

A utilização da DEA poderá ser complementada por táticas como a normalização das informações, a ampliação da base de comparação (com representatividade nacionalmente) e a aplicação de métodos complementares como *bootstrap*² e DEA estocástico³. A relação mínima entre as variáveis e as DMUs foi mantida, e a avaliação será feita por meio do RStudio⁴, utilizando pacotes específicos, possibilitando a criação de indicadores, gráficos e análises comparativas.

Extração e Interpretação da Matriz de Lambdas e da DMU Virtual

Além do cálculo dos escores de eficiência, o presente estudo extraiu dois subprodutos do modelo DEA para aprofundar o diagnóstico da eficiência da UFAM: a matriz de lambdas e a DMU Virtual.

Procedimento de extração: utilizando o pacote *deaR* (Coll-Serrano; Bolós; Suárez, 2022) no ambiente R, após a execução do modelo *model_basic()* com orientação a *outputs* (*orientation*="oo") e retornos variáveis de escala (*rts*="vrs"), aplicou-se a função *lambdas()* ao objeto de resultados. Essa função retorna uma matriz de dimensão $n \times n$ (onde n é o número de DMUs), na qual cada célula λ_{ji} representa o peso da DMU j (*benchmark*) na composição da fronteira de referência para a DMU i (avaliada).

² *Bootstrap* DEA é uma técnica estatística utilizada em conjunto com a Análise Envolvente de Dados para validar e obter inferências estatísticas mais robustas sobre os resultados de eficiência.

³ DEA estocástico é uma extensão do método DEA, utilizada para medir a eficiência de unidades produtivas quando os dados de inputs e outputs estão sujeitos a incertezas, erros de medição ou flutuações aleatórias, em vez de serem valores fixos e determinísticos.

⁴ RStudio é um ambiente de desenvolvimento integrado (IDE) de código aberto para a linguagem de programação R.

Interpretação dos λ para a UFAM (ANEXO 1 – B. MATRIZ DE LAMBDA): a linha correspondente à UFAM na matriz de lambdas contém os coeficientes λ para cada uma das 29 IES da amostra. Valores de $\lambda > 0,001$ foram considerados contribuições efetivas para a composição da DMU Virtual. A soma dos λ é igual a 1 (devido ao modelo VRS), o que permite interpretá-los como percentuais de participação de cada *benchmark* na referência composta.

Cálculo da DMU Virtual e das metas: a DMU Virtual da UFAM foi construída como a combinação linear ponderada dos *outputs* das DMUs com $\lambda > 0$. Formalmente:

$$Y_{Virtual}^{UFAM} = \sum_{j \in E} \lambda_j \cdot Y_j^{real}$$

onde E é o conjunto de DMUs com $\lambda > 0,001$. Os valores observados da UFAM foram comparados com os valores da DMU Virtual para quantificar as folgas (*slacks*) de *outputs*, conforme a seguinte relação (modelo orientado a *outputs*):

$$\text{Folga}(\%) = \left(\frac{Y_{Virtual} - Y_{Real}}{Y_{Real}} \right) \times 100$$

Ressalva metodológica: os valores das metas (ex: 121 artigos e 51 patentes) não configuram prescrições operacionais ou metas institucionais, mas sim referências para investigação qualitativa. Conforme discutido no Referencial Teórico (Seção 2.7), a literatura DEA recomenda que tais targets sejam tratados como referências para investigação qualitativa, pontos de partida para perguntas sobre práticas institucionais, não como alvos numéricos de gestão (Thanassoulis; Portela; Despić, 2008). Esta ressalva foi incorporada à interpretação dos resultados apresentada no Capítulo 2 (Seção 2.6) e na Síntese dos Achados (Seção 5.1).

Inteligência Artificial (NLP/ML): para a avaliação da qualidade e potencial preditivo do conhecimento científico

O NLP será empregado para analisar teses e dissertações do PPGBIOTEC/UFAM, buscando identificar objetos de pesquisas que possam se tornar

patenteáveis e que possam também fornecer suporte à formulação de políticas institucionais de inovação.

Figura 3 - Diagrama NLP



Fonte: Elaborado pelo autor (2025).

O ML permitirá explorar a antecipação de áreas potenciais para inovação e para apoiar a tomada de decisões estratégicas; nesse caso, os modelos serão alimentados com dados institucionais e bibliográficos para descobrir oportunidades subexploradas e aprimorar a alocação de recursos.

Figura 4 - Diagrama ML



Fonte: Elaborado pelo autor (2025).

Limitações e Desafios da IA (NLP/ML).

A aplicação de IA exige atenção a vieses, representatividade e qualidade dos dados, mitigados aqui por validação contextual e métricas de equidade.

Viés de representação e sobreajuste (overfitting).

Modelos conseguem reproduzir preconceitos presentes nos dados de treinamento, ignorando categorias minoritárias ou criando previsões tendenciosas. A ocorrência de *overfitting* requer aplicar métodos de regularização e validação cruzada para assegurar a generalização, além de serem necessárias abordagens como aumento de dados, modificações em *embeddings* e a participação das comunidades impactadas (Shaw; Goïta; Germain, 2022).

2.3. DETALHAMENTO DA COLETA E DO TRATAMENTO DE DADOS

A coleta e o tratamento dos dados são realizados em duas etapas sucessivas (DEA e IA/NLP/ML), essenciais para o *Framework* Sequencial e para garantir a robustez e a comparabilidade dos dados.

Tratamento de Dados para a Análise de Eficiência (DEA)

A etapa de DEA foca na criação de uma Matriz de Decisão que quantifica o desempenho da transferência tecnológica nas Unidades de Tomada de Decisão (DMUs).

Os dados de *Inputs* e *Outputs* para a análise de eficiência das IES em relação à Transferência Tecnológica em Biotecnologia podem considerar, mas não se restringem a: RH, Produção Científica, Bolsas, Projetos, Patentes, etc., que podem ser coletados das bases oficiais, como: CAPES, SCOPUS, PATENTSCOPE, INEP, RUF, entre outras. Para manter consistência terminológica, as variáveis utilizadas no modelo DEA são tratadas exclusivamente como insumos (*inputs*) e produtos (*outputs*), evitando interpretações ambíguas como 'infraestrutura' ou 'recursos', que não correspondem à modelagem adotada.

Harmonização e Limpeza

Para garantir a comparabilidade temporal, todos os dados devem ser consolidados por faixa de ano e harmonizados sob a mesma definição de DMU (Instituições de Ensino Superior com Programas de Pós-Graduação em

Biotecnologia). Dados faltantes ou inconsistentes nas séries históricas são tratados via imputação linear, ou, em casos extremos, exclusão da DMU em questão, conforme técnicas estatísticas.

As variáveis foram submetidas a harmonização logarítmica, procedimento que reduz assimetrias, melhora a comparabilidade entre unidades e evita distorções no cálculo dos escores de eficiência.

Tratamento de Dados para a Prospeção com Inteligência Artificial (IA/NLP e ML)

A etapa de IA/NLP foca na mineração e classificação do conhecimento produzido pelo PPGBIOTEC/UFAM, utilizando os textos integrais de teses e dissertações (TEDE UFAM) como *corpus*.

O *corpus* é composto pelos textos integrais em PDF (*Portable Document Format*) integrais dos trabalhos do PPGBIOTEC/UFAM defendidos em determinado período, coletados via Repositório Institucional.

O pré-processamento é essencial para a acurácia dos modelos de IA/NLP e é realizado por meio de *scripts* no RStudio, conforme etapas detalhadas:

- Conversão dos textos brutos em unidades léxicas (tokens), remoção de caracteres especiais, números e pontuações.
- Remoção de *stopwords* (palavras de alta frequência, como artigos e preposições) e aplicação da técnica de *stemming* (redução das palavras ao seu radical) para otimizar o treinamento do modelo e reduzir a dimensionalidade dos dados.

Os textos processados são transformados em um formato vetorizado (*Term Frequency-Inverse Document Frequency* - TF-IDF) e utilizados para alimentar os modelos de *Machine Learning* para a classificação de TRL e Códigos CIP relacionados à biotecnologia, como os Biotec-BR.

2.4. ESTUDOS DE CASO

A pesquisa utiliza um estudo de caso único, UFAM e PPGBIOTEC, como sua principal abordagem, possibilitando uma análise detalhada de processos institucionais, políticas públicas e práticas inovadoras na biotecnologia. Essa abordagem é apoiada por vários autores, incluindo Silva, Oliveira e Silva (2021), que

argumentam seu valor em contextos específicos e regionais. De acordo com Yin (2014), essa metodologia é adequada para pesquisas exploratórias, fornecendo uma perspectiva contextual e minuciosa dos fenômenos em seu ambiente original.

Ademais, Baxter e Jack (2015) observam que estudos de caso cuidadosamente elaborados possibilitam a fusão de abordagens quantitativas e qualitativas, ajudando a analisar interações complexas. A fundamentação teórica consolidada destaca a relevância dessa escolha metodológica para compreender os obstáculos e as oportunidades da inovação na realidade amazônica.

2.5. CONSOLIDAÇÃO DAS ETAPAS METODOLÓGICAS

A investigação utiliza uma metodologia combinada, de acordo com Silva *et al.* (2025), unindo técnicas qualitativas e quantitativas para examinar a transferência de tecnologia nas Instituições de Ensino Superior.

Etapa Qualitativa, será conduzida uma análise documental metódica de documentos, arquivos de propriedade intelectual e iniciativas de extensão, fundamentada nos critérios estabelecidos por Valle e Ferreira (2025). A técnica de análise de conteúdo, segundo Cardoso, Oliveira e Ghelli (2021) e Paiva, Oliveira e Hillesheim (2021), será empregada para interpretar significados e unidades discursivas relevantes. A condução será apoiada por metodologias ágeis, que permitem ciclos iterativos e controle analítico (Coutinho; Andrade; Santana, 2024).

Etapa Quantitativa, a eficiência institucional será mensurada via DEA, conforme Teixeira *et al.* (2021); Khezrimotlagh *et al.* (2019) e Ferreira (2020), utilizando o *software* R. Essa abordagem permitirá comparar a UFAM com outras IES, conforme diretrizes de Loveth (2025).

Avaliação Técnico-Científica, utilização de abordagens bibliométricas e métodos de Processamento de Linguagem Natural para identificar campos com maior capacidade de inovação e possibilidade de registro de patentes (Kaur; Malik, 2020).

Machine Learning, modelos preditivos serão utilizados para antecipar regiões com alto potencial biotecnológico, segundo proposta de Chang (2023).

Visualização dos Dados, painéis interativos serão construídos com base nos princípios de design informacional (Aveni; Carvalho, 2017), estimulando uma cultura de inovação aberta nas universidades (Marques *et al.*, 2021).

Análise de Cenário Regional, será conduzida uma articulação dos resultados com o contexto socioambiental amazônico, permitindo identificar implicações institucionais e potenciais caminhos de aprimoramento.

2.6. INTEGRAÇÃO E ANÁLISE DOS RESULTADOS

Os resultados são integrados em duas fases, seguindo o modelo sequencial proposto. Ao unir o diagnóstico de eficiência (DEA) com a prospecção tecnológica (IA/NLP/TRL), é possível associar o desempenho de uma instituição à sua capacidade de criar ativos científicos que possam evoluir tecnologicamente. Assim, é possível identificar as habilidades internas, os tipos de produção científica e os sinais de desenvolvimento tecnológico nos documentos que foram analisados.

Os resultados da DEA, a Eficiência Técnica Relativa do PPGBIOTEC, as DMUs de referência e as características das unidades eficientes e ineficientes são discutidos na primeira etapa. Na segunda fase, esses resultados são comparados com os indicadores gerados pelos modelos de IA/NLP/ML, em particular no que diz respeito ao potencial de patenteamento e às classificações de maturidade tecnológica (TRL).

A análise cruzada possibilita identificar diferentes configurações de atuação institucional, ressaltando aquelas áreas em que a produção científica se alinha mais ao desenvolvimento tecnológico e aquelas em que os esforços continuam majoritariamente voltados para a pesquisa básica. Entre os padrões observados, destacam-se:

- Foco em insumos de pesquisa: quando o potencial de patenteabilidade e o TRL são baixos, ou seja, quando atividades científicas em estágios iniciais de desenvolvimento são mais comuns que inovações tecnológicas voltadas ao mercado.
- Foco na conversão tecnológica: situações em que a capacidade inovadora é alta, mas a eficiência comparativa na transformação dessa capacidade em ativos de propriedade intelectual é baixa, indicando a presença de entraves institucionais na conversão.

Os resultados integrados não têm caráter prescritivo. O objetivo é proporcionar uma análise das interconexões entre eficiência institucional, geração de conhecimento científico e maturidade tecnológica, visando elucidar as dinâmicas que moldam a transferência tecnológica em biotecnologia na Amazônia.

2.7. LÓCUS DA PESQUISA: UNIVERSIDADE FEDERAL DO AMAZONAS (UFAM)

O lócus desta pesquisa compreende três instâncias centrais da UFAM: a própria instituição enquanto agente público de ciência e tecnologia; o PPGBIOTEC, como ambiente de produção científica especializada; e a PROTEC, responsável pela gestão da política institucional de inovação, propriedade intelectual e transferência de tecnologia. A articulação entre esses níveis institucionais configura o ecossistema no qual se insere este estudo, orientado à compreensão de capacidades, fluxos e potenciais de inovação no contexto amazônico.

A UFAM, sucessora da Escola Universitária Livre de Manáos (1909), é a mais antiga instituição de ensino superior do Brasil e exerce papel estratégico na formação de recursos humanos e na produção científica na região Norte. Reconhecida como Instituição Científica, Tecnológica e de Inovação (ICT), possui políticas próprias de inovação regulamentadas pelas Resoluções CONSUNI nº 009/2011 e nº 011/2023, que estabelecem diretrizes para proteção intelectual, pesquisa aplicada e interação com o setor produtivo. A PROTEC integra essa estrutura, desde 2011, atuando como Núcleo de Inovação Tecnológica (NIT), no que se refere à gestão da Política de Inovação e da Propriedade Intelectual da UFAM.

O PPGBIOTEC, criado em 2001, consolidou-se como um dos principais polos de formação avançada em biotecnologia na Amazônia. Com áreas de concentração que abrangem biotecnologias agroflorestais, biotecnologias para a saúde e gestão da inovação, o programa reúne pesquisadores de instituições como UFAM, INPA, FIOCRUZ/AM, Embrapa Amazônia Ocidental, FMT-IMT e SUFRAMA. Sua produção científica, registrada no repositório TEDE/UFAM, constitui o corpus analisado nesta pesquisa por meio de técnicas de Processamento de Linguagem Natural e *Machine Learning*.

A escolha do PPGBIOTEC como lócus específico decorre de três fatores principais: (i) sua relevância regional e nacional na formação de recursos humanos em biotecnologia; (ii) o volume expressivo de teses e dissertações com potencial de inovação; e (iii) a necessidade institucional de fortalecer a articulação entre pesquisa acadêmica e transferência tecnológica, conforme previsto nas políticas de inovação da UFAM.

A literatura institucional da UFAM registra marcos relevantes na consolidação da política de inovação. Chaves (2012) analisa a criação do Parque Científico e Tecnológico para Inclusão Social (PCTIS), situando-o como iniciativa alinhada às especificidades amazônicas. Em continuidade, Chaves (2018) revisita aspectos normativos e administrativos da gestão da inovação na UFAM, descrevendo a evolução de estruturas internas relacionadas à propriedade intelectual e à articulação institucional.

Esses estudos não se contrapõem à trajetória institucional descrita por Chaves, mas a complementam, ao evidenciar que o fortalecimento da política de inovação envolve tanto avanços normativos e estruturais quanto a necessidade de aprimorar mecanismos operacionais, analíticos e digitais. Nesse sentido, o *Framework* Sequencial desenvolvido nesta tese, composto por análises de eficiência (DEA), potencial inventivo e maturidade tecnológica (NLP/TRL) e predição de *hotspots* (ML), foi aplicado ao caso do PPGBIOTEC/UFAM para indicar, ao nível conceitual, como diferentes métodos de inteligência de dados podem ser articulados para revelar padrões, lacunas e oportunidades no ecossistema institucional.

O *Dashboard* Interativo apresentado no Capítulo 5 deve ser compreendido como protótipo conceitual, destinado exclusivamente à integração visual dos resultados analíticos. Sua descrição não constitui proposta de implementação, e qualquer eventual adoção futura dependeria de decisões institucionais próprias, prioridades estratégicas e mecanismos de fomento externos, aspectos que não integram o escopo desta tese. A menção a tais possibilidades, quando presente, tem caráter meramente contextual, refletindo discussões recorrentes na literatura sobre modernização da gestão da inovação em ICTs públicas.

Considerando esse escopo metodológico e a natureza integrada das etapas analíticas, torna-se necessário explicitar também as condições de uso das técnicas de IA empregadas nesta pesquisa.

Governança e a responsabilidade se tornaram cada vez mais importantes devido ao aumento considerável do uso da Inteligência Artificial nos âmbitos acadêmico e de pesquisa. Tal fato se manifesta em esforços como o *AI Act* da União Europeia (European Commission, 2024) e o novo *Framework* da OCDE para categorizar e administrar sistemas de inteligência artificial (OCDE, 2022), ao passo que a norma ISO/IEC 42001:2023 (ISO, 2023) oferece diretrizes voltadas para a gestão de sistemas de IA, enfatizando a importância da rastreabilidade, mitigação de

vieses, documentação apropriada e transparência nos métodos utilizados. A portaria CNPq nº 2.664/2026 (MCTI, 2026) no Brasil tem o mesmo intuito de estabelecer diretrizes para a aplicação ética e passível de verificação dessas tecnologias no campo científico.

Segundo as diretrizes estabelecidas por essas normas internacionais e nacionais, o presente trabalho descreve as etapas adotadas. Aborda, igualmente, as dificuldades enfrentadas e as estratégias empregadas para validar as fases de processamento de linguagem natural e aprendizado de máquinas. Essas ferramentas foram empregadas como métodos para lidar com grandes volumes de dados ou dados complexos. A interpretação conclusiva cabe ao ser humano, assegurando que a forma de atuação seja transparente e conforme as diretrizes previamente indicadas.

Portanto, a aplicação da inteligência artificial nesta investigação não está restrita a utilizar métodos computacionais. É igualmente imprescindível esclarecer quais são as condições que possibilitam que esses métodos proporcionem resultados passíveis de verificação. As fases de NLP e ML foram realizadas com diligência, a fim de assegurar que seja viável monitorar, validar e preservar a uniformidade dos procedimentos. Portanto, a aplicação dessas técnicas está relacionada ao planejamento da pesquisa. Isso contribui para compreender as potencialidades e as restrições dos processos de busca tecnológica examinados.

3. RESULTADOS DA PESQUISA

A estrutura desenvolvida até aqui estabelece as bases conceituais, metodológicas e institucionais necessárias para a apresentação dos resultados empíricos desta tese. A fundamentação teórica evidenciou não apenas a necessidade de aprimorar a capacidade da UFAM de transformar conhecimento científico em inovação aplicável, mas também que parte significativa desse potencial já se encontra presente em sua própria produção acadêmica, ainda pouco explorada ou sistematicamente identificada. Essa capacidade, além de estratégica, constitui um dever legal das Instituições Científicas, Tecnológicas e de Inovação.

Tais deveres estão expressos na Política de Desenvolvimento da Biotecnologia (Decreto 6.041/2007), na Lei de Inovação (Lei 10.973/2004) e no Marco Legal de CT&I (Lei 13.243/2016), que determinam que ICTs públicas devem:

- (i) gerir adequadamente sua propriedade intelectual;
- (ii) promover a transferência de tecnologia;
- (iii) estruturar e manter um Núcleo de Inovação Tecnológica funcional;
- (iv) adotar mecanismos de prospecção, avaliação e valoração de ativos tecnológicos.

No contexto das ICTs públicas brasileiras, é comum que estruturas designadas para exercer funções de Núcleos de Inovação Tecnológica concentrem sua atuação em atividades jurídico-administrativas. Na UFAM, a PROTEC foi designada para desempenhar essas atribuições, desenvolvendo ações de proteção da propriedade intelectual, ainda que com foco predominante nos procedimentos formais previstos em lei.

Essa configuração, amplamente observada no sistema nacional de CT&I, contribui para a existência de lacunas entre as atribuições legais dos NITs e as práticas efetivamente implementadas. Os resultados desta tese sugerem que tais lacunas ajudam a explicar a persistência de gargalos na prospecção tecnológica, na gestão ativa de portfólio e na identificação de ativos estratégicos, elementos que reforçam a relevância de abordagens baseadas em inteligência de dados para apoiar o aprimoramento institucional.

O referencial teórico consolidou os elementos que sustentam essa trajetória, abordando desde as limitações das métricas tradicionais de transferência tecnológica até o papel da inteligência de dados na avaliação de maturidade tecnológica e na

identificação de ativos biotecnológicos. A metodologia, organizada por meio do *Framework* Sequencial, operacionalizou esses fundamentos ao integrar a Análise Envolvória de Dados, técnicas de Processamento de Linguagem Natural e modelos de *Machine Learning* aplicados ao acervo de teses e dissertações do PPGBIOTEC.

Com a execução das etapas iniciais do *Framework*, torna-se possível apresentar os primeiros resultados empíricos, distribuídos nos capítulos subsequentes:

- Capítulo 1 apresenta o panorama nacional de patentes em biotecnologia (2014-2023), contextualizando o posicionamento da UFAM e evidenciando a distância entre produção científica e proteção intelectual.
- Capítulo 2 expõe o diagnóstico de eficiência institucional por meio da DEA, comparando o desempenho inovador da UFAM com outras universidades públicas brasileiras.
- Capítulo 3 apresenta os resultados da classificação automática de potencial de patenteamento (CIP) e maturidade tecnológica (TRL) no PPGBIOTEC, revelando ativos latentes e padrões temáticos relevantes.
- Capítulo 4 introduz o modelo preditivo para identificação de *hotspots* de bioprospecção na Amazônia, ampliando a dimensão estratégica da análise.
- Capítulo 5 integra os achados dos capítulos anteriores, apresentando o *Framework* Sequencial como modelo analítico e o *Dashboard* Interativo como síntese visual e operacional das evidências produzidas.

Os resultados obtidos até este ponto cumprem duas funções complementares:

- I. indicam a viabilidade analítica do *Framework* Sequencial, validando o uso combinado de DEA, NLP e ML para diagnóstico e prospecção tecnológica;
- II. oferecem uma visão preliminar de padrões, tendências e potenciais, que serão aprofundados nos capítulos empíricos e integrados na síntese final da tese.
- III. Assim, esta seção marca a transição entre a construção metodológica e a análise dos achados, preparando o terreno para a compreensão integrada do potencial inovador da UFAM e de sua inserção estratégica na bioeconomia amazônica.

A partir desse arcabouço, os capítulos seguintes apresentam os resultados empíricos organizados em forma de artigos, cada um aprofundando uma dimensão específica da análise.

CAPÍTULO 1 - TRANSFERÊNCIA TECNOLÓGICA: FATOS E PERCEPÇÕES SOBRE PATENTES EM BIOTECNOLOGIA NO BRASIL DE 2014 A 2023⁵

1.1. INTRODUÇÃO: O CONTEXTO DA INOVAÇÃO E A NECESSIDADE DE FORTALECER A TRANSFERÊNCIA TECNOLÓGICA NA AMAZÔNIA

A biotecnologia, nas últimas décadas, consolidou-se como uma força transformadora, moldando o fundamento do conhecimento científico e o progresso econômico e tecnológico de inúmeras nações, sendo ela mesma uma área do conhecimento que combina biologia e tecnologia, que deve poder revolucionar setores como a medicina, a agricultura e a indústria, trazendo soluções inovadoras para problemas globais, como a segurança alimentar e a cura de doenças. Quando o mundo enfrenta problemas complexos, a biotecnologia se destaca como uma importante ferramenta para grandes e sustentáveis progressos (Rosa; Sabonaro, 2019; Silva, 2024).

Assim, as patentes têm uma importância crítica nesse contexto, não apenas como uma maneira de proteger inovações, mas também podem ser consideradas um sinal claro de progresso e criatividade, em outras palavras, um facilitador de inovações futuras. No entanto, não é suficiente, mesmo quando se trata de patentes, que as pessoas pensem no valor intrínseco da criação de novas tecnologias. É necessário promover uma comunicação ativa e a transferência de tecnologia entre universidades e setores privados, para que essa ideia se torne um produto e possa ser disponibilizada à sociedade, e esse incentivo, em última instância, permite que empresas e países prosperem no mercado (Alves *et al.*, 2018).

Este artigo visa estudar o estado das patentes na biotecnologia no Brasil, utilizando o conjunto de dados denominado Biotec-BR. A pesquisa foi feita com base nas informações coletadas do LATIPAT-ESPACENET (LATIPAT, 2017), plataforma que disponibiliza o acesso aos documentos de patentes. A análise dos dados deste estudo busca não somente entender as tendências atuais, mas também gerar

⁵ O presente Capítulo é uma versão expandida e reformatada do artigo intitulado 'Transferência tecnológica: fatos e percepções sobre patentes em biotecnologia no Brasil de 2014 a 2023', publicado na Revista Observatório de la Economía Latinoamericana, Vol. 23, N. 2, p. 01–21, 2025, DOI: 10.55905/oelv23n2-072, (KUME; OLIVEIRA JÚNIOR, 2025). O texto base das seções 1.2, 1.3 e 1.4 é reproduzido integralmente, conforme normas de autocitação e tese-artigo.

conhecimento que poderá ser aplicado em futuras pesquisas, contribuindo para o avanço do campo e promovendo a elaboração de estratégias para um sistema biotecnológico mais efetivo.

A metodologia adotada neste estudo baseou-se na realização de buscas direcionadas por patentes registradas no período de 2014 a 2023, com foco específico em inovações no campo da biotecnologia. Os dados coletados foram organizados e analisados por meio de uma abordagem qualiquantitativa, possibilitando uma compreensão abrangente sobre o panorama das patentes no Brasil ao longo do período investigado. A identificação de oportunidades estratégicas para inovações no setor foi evidenciada pela relevância da necessidade de um sistema eficiente de proteção da propriedade intelectual como elemento impulsionador do desenvolvimento da biotecnologia no país.

1.2. FUNDAMENTOS CONCEITUAIS: A PROPRIEDADE INTELECTUAL, OS FATOS DA BIOTECNOLOGIA E A PERCEPÇÃO DO CONHECIMENTO

Patentes de Biotecnologia

As patentes na biotecnologia, incentivam o desenvolvimento de novas tecnologias, criando um ambiente favorável para investimentos em pesquisa e garantindo a proteção de inovações. O período de exclusividade concedido aos inventores mediante a solicitação de uma patente possibilita que esses profissionais colham os frutos do seu trabalho criativo, o que, por sua vez, contribui para o avanço científico e médico. Exemplos como as terapias genéticas baseadas em genes ilustram como os direitos de propriedade intelectual podem levar a avanços médicos significativos, beneficiando empresas e pacientes (Palavecino, 2021; Pinto; de Matos, 2021).

O processo de obtenção de uma patente é repleto de detalhes técnicos e legais que podem ser desafiadores até para os mais experientes. Aliado a isso, a burocracia e a morosidade na concessão de patentes, erguem obstáculos que dificultam o ritmo da inovação, prejudicando tanto pesquisadores quanto empresas (Alba; Sugimoto; Morioka, 2021). No entanto, há exemplos que ilustram como, esse sistema complexo pode trazer grandes resultados. Um caso claro disso é o da soja Roundup Ready,

patenteada pela Monsanto, que transformou a agricultura ao oferecer resistência a herbicidas, provando o poder das patentes de impulsionar mudanças significativas em diversos setores (Ferreira; Terra; Gregori, 2020).

A importância das patentes também merece atenção, em instituições de pesquisa, como a Fundação Oswaldo Cruz (FIOCRUZ). Um estudo realizado com teses e dissertações de lá revelou que uma parcela diminuta dos trabalhos mencionava patentes. Isso indica que, apesar da relevância desse tema para a biotecnologia, muitos pesquisadores ainda não consideram a propriedade intelectual como parte integrante de suas pesquisas.

A formação acadêmica, especialmente em áreas que lidam com biotecnologia para a saúde, precisa incorporar mais profundamente esse aspecto, ajudando a preparar melhor os futuros cientistas (Pimenta, 2017).

Instituições brasileiras demonstram um crescente envolvimento com o patenteamento, principalmente no setor de fármacos e cosméticos, aproveitando o potencial de suas riquezas naturais. Essa movimentação não só reforça o papel das patentes no desenvolvimento de soluções inovadoras, como também representa um avanço no reconhecimento e valorização do conhecimento local, com implicações positivas para a sociedade na totalidade (Rabelo *et al.*, 2023).

Portanto, o sistema de patentes é necessário para proteger as inovações e garantir que os inventores possam usufruir comercialmente dos frutos de suas criações, criando um ciclo de desenvolvimento que beneficia a sociedade. No entanto, é necessário equilibrar a proteção da inovação com o acesso às tecnologias, um desafio que é particularmente evidente em setores como o da biotecnologia, em que grandes corporações detêm frequentemente o controle das patentes, limitando o acesso de pequenos produtores e consumidores a tecnologias essenciais. A busca por esse equilíbrio continua sendo um tema central no debate sobre propriedade intelectual na biotecnologia (Amaral *et al.*, 2020).

Panorama de Estudos Anteriores

A pesquisa sobre patentes em biotecnologia no Brasil revela avanços, mas também, grandes desafios. Entre os obstáculos mais destacados, estão a complexidade do sistema de patentes, a burocracia e a necessidade de investimentos substanciais em pesquisa. A falta de dados sistematizados e a dificuldade de acesso

a informações sobre os requerentes também contribuem para limitar a eficácia do sistema, dificultando o progresso do setor (Bochi; Moura, 2020). Embora o Brasil tenha experimentado um aumento no número de patentes entre 1978 e 2018, esse crescimento não é suficiente para consolidar o país como um líder global em inovação biotecnológica. As dificuldades financeiras e a escassez de infraestrutura, aliadas à demora na concessão de patentes, são fatores críticos nesse contexto (Fernandes; Gadelha; Maldonado, 2021).

No entanto, a análise das patentes do Brasil também aponta algumas conquistas notáveis. A pesquisa sobre cultivares de soja, por exemplo, evidenciou o impacto da biotecnologia na agricultura, com inovações que aumentam a produtividade e melhoram a resistência a doenças (Vellani Júnior *et al.*, 2022). Casos como o da Universidade Estadual do Ceará, que registrou uma patente para um fungicida derivado da casca de castanha de caju, mostram o esforço contínuo em pesquisa e desenvolvimento, apesar das limitações financeiras. A trajetória dessa patente reflete a luta por inovação e reconhecimento no cenário científico, sendo um exemplo das dificuldades e sucessos enfrentados pelos pesquisadores brasileiros (Ipiranga; Chaym; Sousa, 2016).

Além disso, a colaboração entre instituições é um ponto positivo no avanço da biotecnologia no Brasil. A Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), por exemplo, se destaca pela realização de patentes conjuntas, demonstrando o potencial de avanços no setor (Camara *et al.*, 2015). As patentes permitem que empresas protejam seus investimentos e impulsionem o progresso tecnológico, sendo necessárias para proteger as inovações, elas também promovem a troca de conhecimento entre a pesquisa e a indústria, como ocorre com a FIOCRUZ, contribuindo com o desenvolvimento de soluções sustentáveis e inovadoras, principalmente para a produção de proteínas alternativas a partir de processos fermentativos (Ribeiro *et al.*, 2023).

Por outro lado, o Brasil, além da competição internacional aliada à biopirataria, continua dependente de inovações estrangeiras. Entre 1978 e 2012, o país registrou 473 patentes, das quais somente 130 foram depositadas originalmente no Brasil, com 111 provenientes de residentes internos. Além disso, a interação entre os agentes de inovação ainda é insuficiente, com somente 13,51% das patentes resultando de colaborações (Camara *et al.*, 2015). A baixa taxa de inovação na indústria farmacêutica brasileira, refletida na escassez de patentes, é outro fator preocupante,

com o país representando somente 0,06% das patentes globais nesse setor em 2015. A colaboração com empresas internacionais, como nas vacinas contra a COVID-19, ilustra tanto as capacidades quanto as dificuldades do setor (Fernandes; Gadelha; Maldonado, 2021).

Por fim, o crescente interesse e potencial no setor de biotecnologia no Brasil são evidentes, especialmente em áreas como fitoterápicos e cosméticos. As patentes relacionadas a extratos de algas para uso cosmético e tecnologias limpas, refletem uma busca por soluções sustentáveis e uma maior utilização de recursos biológicos para inovação (Cardoso *et al.*, 2022). No entanto, para que o Brasil se torne mais competitivo globalmente, será necessário um ambiente regulatório mais favorável e investimentos mais robustos em pesquisa e desenvolvimento.

Classificação Internacional De Patentes (CIP) De Biotecnologia: Biotec-BR

Certamente, a biotecnologia é um dos campos cada vez mais em evidência, no que diz respeito ao desenvolvimento econômico e social, principalmente em países com riqueza de biodiversidade, como o Brasil. A proteção das invenções por patentes é um aspecto a considerar, visto que esta protege os pesquisadores e inventores, garantindo que usufruam dos benefícios que adquirirem de suas invenções, incentivando a continuidade da pesquisa e do desenvolvimento (Villela *et al.*, 2022).

No Brasil, a análise das patentes de biotecnologia, principalmente entre 2012 e 2016, revelou a necessidade de uma classificação mais detalhada das inovações. Os resultados indicaram que CIPs específicos poderiam levar à recuperação de documentos relevantes abrangendo tecnologias relacionadas a áreas como meio ambiente e agroindústria, essa estratégia não somente amplia o banco de dados utilizado, mas também apresenta uma imagem mais clara da capacidade de inovação do país.

O Instituto Nacional da Propriedade Industrial (INPI) tem sido um grande protagonista nesse cenário, promovendo o uso das CIPs como ferramenta de prospecção tecnológica no campo da biotecnologia. Com essa iniciativa, o INPI se posiciona como um agente importante na promoção do desenvolvimento tecnológico do país e na criação de políticas públicas que valorizem a inovação e incentivem o uso sustentável dos recursos naturais brasileiros (de Medeiros; Pelaez, 2021).

As CIPs são importantes para a organização das patentes e, por conseguinte, das inovações biotecnológicas. A proposta de elencar um conjunto relacionado à biotecnologia, denominada Biotec-BR, foi conduzida para atender às necessidades específicas do setor no Brasil, permitindo uma identificação mais precisa de patentes que poderiam ser negligenciadas em classificações não pertinentes. Essa abordagem facilita a análise do cenário de patentes, possibilitando a identificação de tendências e oportunidades de inovação essenciais para o crescimento do setor (Weid *et al.*, 2018).

1.3. ESTRATÉGIA DE MAPEAMENTO: FONTES, COLETA E DELIMITAÇÃO TEMPORAL PARA O ESTUDO RETROSPECTIVO (2014-2023)

A metodologia deste estudo consistiu em uma busca direcionada de patentes relacionadas à biotecnologia registradas entre 2014 e 2023. O levantamento foi realizado por meio do LATIPAT-ESPACENET, uma interface especializada em documentos de patentes que oferece acesso gratuito a informações do INPI e de outras bases da América Latina e Espanha. A pesquisa utilizou o conjunto de CIPs conhecido como Biotec-BR, que agrupa códigos específicos da biotecnologia. Esse enfoque possibilitou o mapeamento de inovações relevantes, destacando o perfil dos requerentes, especialmente Instituição de Ensino Superior (IES), e as regiões mais ativas no setor, ainda que limitado pelo recorte e pela abrangência da ferramenta empregada.

Plataforma de Busca de Patentes - LATIPAT-ESPACENET

O LATIPAT-ESPACENET, combina a tecnologia do Espacenet com o esforço colaborativo de serviços nacionais de propriedade intelectual de 18 países, incluindo o INPI, sob a coordenação da Organização Mundial da Propriedade Intelectual (OMPI), do Instituto Europeu de Patentes (EPO) e da Oficina Española de Patentes y Marcas (OEPM). Além da pesquisa simplificada por palavras-chave, a ferramenta oferece opções avançadas, permitindo explorar classificações internacionais e consultar documentos em outras coleções globais. Seu acesso intuitivo facilita o uso

por pesquisadores, inventores e empreendedores interessados em inovação e propriedade intelectual na região.

Critérios de Busca de Patentes

A busca foi estruturada com base no conjunto Biotec-BR (Quadro 1), que inclui 80 CIPs relacionadas à biotecnologia. Objetivando reduzir o número de interações com a plataforma LATIPAT-ESPACENET, as classificações foram agrupadas em pares, facilitando o processo, sem perda da abrangência dos resultados. Essa estratégia permitiu uma análise mais ágil, mantendo o foco nas inovações mais relevantes dentro do campo da biotecnologia.

Quadro 1 - Tabela de CIPs do conjunto Biotec-BR

Seção CIP	Classificações propostas no estudo (Biotec-BR)
A	A01H1*, A01H4*, A01H17* A01N63* A21D8/04 A23B4/22, A23B5/16, A23B7/155, A23B9/28 A23C9/12*, A23C13/16, A23C17/02, A23C19/032, A23C21/02 A23F3/10 A23J1/18, A23J3/20 A23L1/105 (transferido para A23L 7/104), A23L1/23 (transferido para A23L 27/24), A23L2/84, A23L3/3571 A61K8/99, A61K35/66, A61K35/68, A61K35/7*, A61K36/06*, A61K36/07*, A61K38*, A61K39*, A61K48*, A61K49/14, A61K49/16, A61K51/08, A61K51/10 A61L15/36, A61L27/38 A62D3/02
B	B01D53/84, B01D59/36 B09C1/10 B82Y5/00
C	C02F3*, C02F9/14, C02F11/02, C02F11/04 C05F11/08 C07K4*, C07K14*, C07K16*, C07K17*, C07K19* C09K8/582 C12C11* C12F* C12G1/022, C12G1/073, C12G3/02 C12M* C12N* C12P* C12Q* C12R* C12S* C40B10*, C40B20*, C40B30*, C40B40*, C40B50*, C40B60*, C40B70*, C40B80*
D	D01C1/04
G	G01N33/53*, G01N33/54*, G01N33/55*, G01N33/56*, G01N33/57*, G01N33/68*

Fonte: Modificado de Weid *et al.* (2018).

As CIPs, organizadas aos pares e separadas pelo operador “OR” conforme o Quadro 2, foram os argumentos utilizados no campo IPC, para a ferramenta de busca das patentes.

Quadro 2 - Combinação e CIPs Agrupadas

Combinação	CIPs	Combinação	CIPs
1	C12N OR C12R	21	C02F11/04 OR C09K8/582
2	C12S OR A61K38	22	B09C1/10 OR A62D3/02
3	A61K39 OR A61K35/66	23	B01D53/84 OR B01D59/36
4	A61K35/68 OR A61K35/7	24	D01C1/04 OR C12M
5	A61K36/06 OR A61K36/07	25	A23C9/12 OR A23C13/16
6	A61L15/36 OR A61K48	26	A23C17/02 OR A23C19/032
7	A61K8/99 OR A61K49/14	27	A23C21/02 OR C12C11
8	A61K49/16 OR A61K51/08	28	C12G1/022 OR C12G1/073
9	A61K51/10 OR A61L27/38	29	C12G3/02 OR A23F3/10
10	C12P OR C12F	30	A23L1/105 OR A23L7/104
11	C07K14 OR C07K16	31	A23L1/23 OR A23L27/24
12	C07K19 OR C07K17	32	A21D8/04 OR A23B4/22
13	C07K4 OR C12Q	33	A23B5/16 OR A23B7/155
14	G01N33/53 OR G01N33/54	34	A23B9/28 OR A23L3/3571
15	G01N33/55 OR G01N33/56	35	A23J1/18 OR A23J3/20
16	G01N33/57 OR G01N33/68	36	A23L2/84 OR B82Y5/00
17	A01H1 OR A01H4	37	C40B10 OR C40B20
18	A01H17 OR A01N63	38	C40B30 OR C40B40
19	C05F11/08 OR C02F3	39	C40B50 OR C40B60
20	C02F9/14 OR C02F11/02	40	C40B70 OR C40B80

Fonte: Elaborado pelo autor (2025).

A utilização da ferramenta LATIPAT-ESPACENET (Figura 1) se deu da seguinte forma: em “Pesquisa Avançada”, com as CIPs combinadas no Quadro 2 no campo “IPC”, com o argumento.

“BR” no campo “Número de pedido”, de modo a obter somente patentes brasileiras; o argumento “[BR]” no campo “Requerente(s)”, de modo a excluir requerentes estrangeiros; e o argumento “2014:2023”, no campo “Data de Publicação”. Cada uma das 40 combinações (uma por vez) foi buscada; os resultados foram adicionados a uma planilha Excel, para posterior análise.

A metodologia previu a utilização das CIPs Biotec-BR para a busca de patentes de biotecnologia e gerou as seguintes considerações: embora a pesquisa tenha recuperado um número significativo de documentos, é passível que algumas inovações relevantes não tenham sido capturadas devido a uma possível limitação do conjunto de CIPs; entretanto, essas observações não devem ser consideradas restrições no sentido crítico, mas sim como oportunidades de refinamento na

metodologia. A análise do patenteamento é um campo dinâmico que pode ser constantemente revisado para melhor refletir o potencial inovativo do Brasil na biotecnologia.

Figura 1 - LATIPAT-ESPACENET

Espacenet

Pesquisa de patentes

Um serviço prestado em cooperação com a IEP

Español Português English

Contacto

Ajuda

« Espacenet Outros serviços online EPO »

Pesquisar

Lista dos resultados

A minha lista de patentes (0)

Histórico

Configurações

Ajuda

Pesquisa inteligente

Pesquisa Avançada

Pesquisa por Classificação

Ajuda rápida

- + Quantos termos posso introduzir por esta campo?
- + Posso efectuar uma pesquisa usando uma combinação de palavras?
- + Como introduzo os termos a partir da descrição ou das reivindicações?
- + Posso usar abreviatura ou símbolos genéricos?
- + Como introduzo números de publicação, pedido, prioridade ou de referência NPL?
- + Como introduzo o nome das pessoas ou das organizações?
- + Qual é a diferença entre a CIP e a CPC?
- + Que formatos posso utilizar para a data de publicação?
- + Como introduzo um intervalo de datas para pesquisa de uma data de publicação?
- + Posso salvar/guardar a minha pesquisa?

Links relacionados

Pesquisa Avançada

Seleccione a coleção em que pretende pesquisar: [1]

LP - Latipat-Espacenet

Digite os seus termos de pesquisa - CTRL-ENTER expande o campo onde você está

Introduza palavras-chave

Palavra(s)-chave no título: [1]

biciçeta and plástico

Palavra(s)-chave no título ou resumo: [1]

rueda

Palavra(s)-chave no texto integral: [1]

rueda

Digite os números com ou sem código de país

Número de publicação: [1]

ES2341329

Número de pedido: [1]

BR

PE20070101320

Número de prioridade: [1]

CU20070000041

Entre uma ou mais datas ou intervalos de datas

Data de publicação: [1]

2014-2023

2003-2004

Digite o nome de uma ou mais pessoas / organizações

Requerente(s): [1]

[BR]

ACINDAR

Inventor(es): [1]

Mendoza

Entre um ou mais símbolos de classificação

CPC [1]

F03G7/10

IPC [1]

C12S OR A5TK38

E01B9/30

Limpar

PESQUISAR

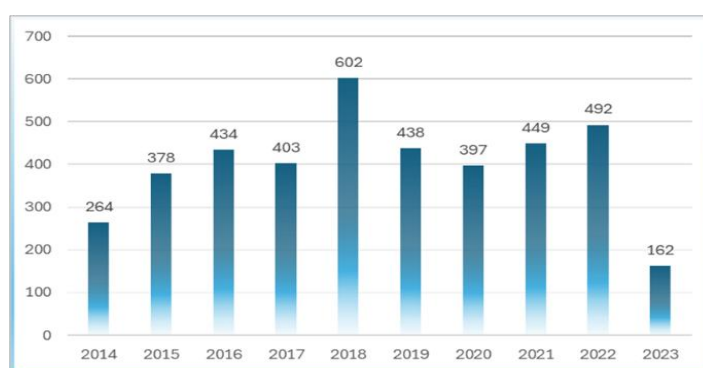
[Mapa do site](#) |
 [Acessibilidade](#) |
 [Notícia legal](#) |
 [Termos de uso](#) |
 Última atualização: 21.11.2017 LP Banco de dados: 6.3.12.00.5; 340

Fonte: Elaborado pelo autor (2025).

1.4. DINÂMICA E ANÁLISE PROSPECTIVA: O PANORAMA DE PATENTES EM BIOTECNOLOGIA E SUAS IMPLICAÇÕES PARA A GESTÃO DE INOVAÇÃO

A busca realizada segundo a metodologia resultou em 4.019 pedidos de patentes de biotecnologia, conforme o conjunto de CIPs definidos pelo Biotec-BR, entre os anos de 2014 e 2023, com média anual de 400 pedidos e desvio padrão de 120, conforme o Gráfico 1.

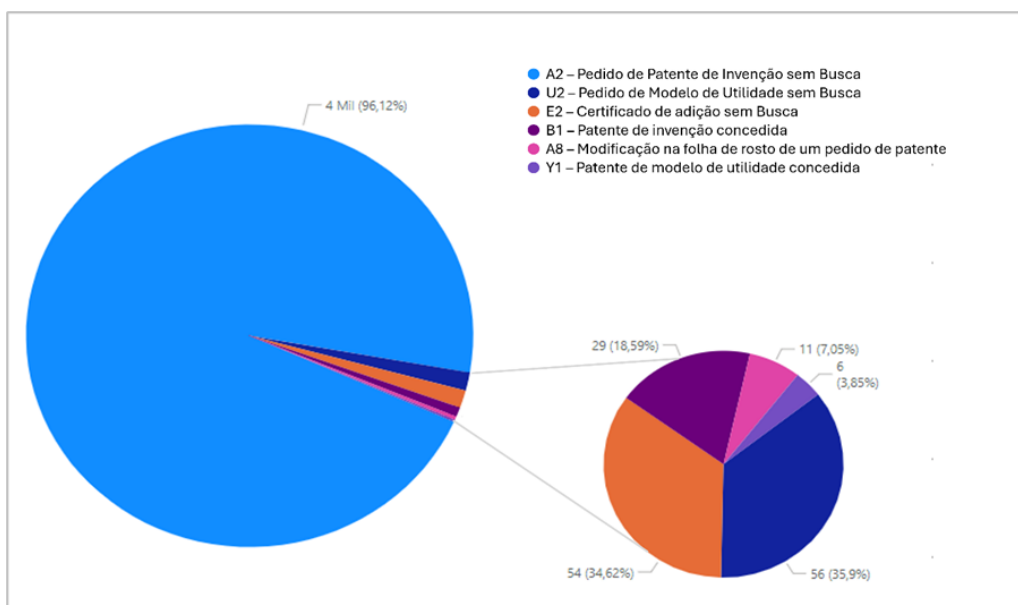
Gráfico 1 - Pedidos de Patentes de Biotecnologia de 2014 a 2023.



Fonte: Elaborado pelo autor (2025).

Dos 4.019 pedidos de patente identificados, somente 29 apresentam condição B1 - Patente de Invenção Concedida, Gráfico 2.

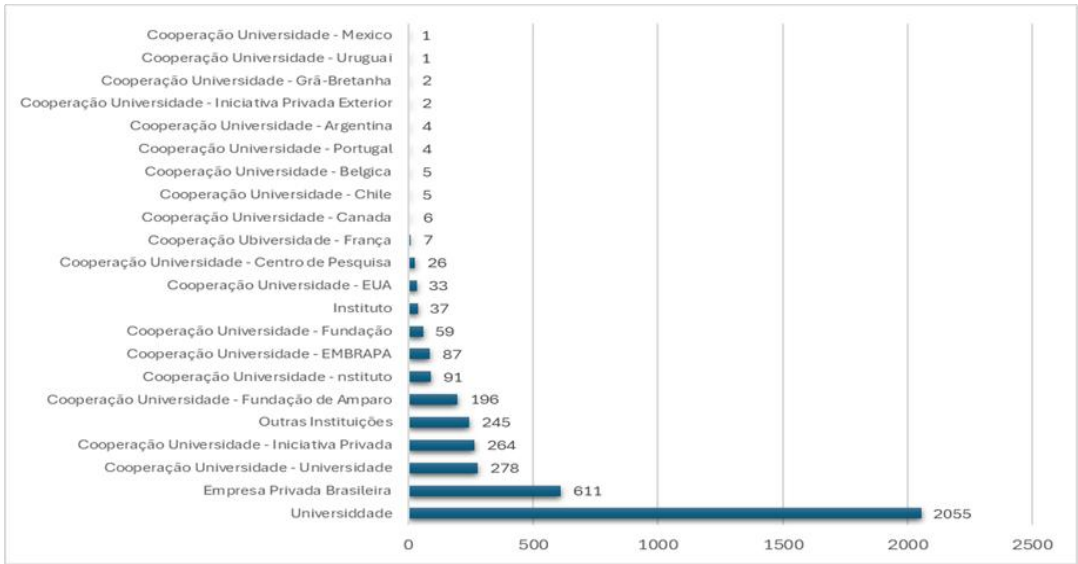
Gráfico 2 - Categorização dos Requerentes de Patentes



Fonte: Elaborado pelo autor (2025).

Foram identificadas 23 categorias de classificação, conforme pode ser observado no Gráfico 3.

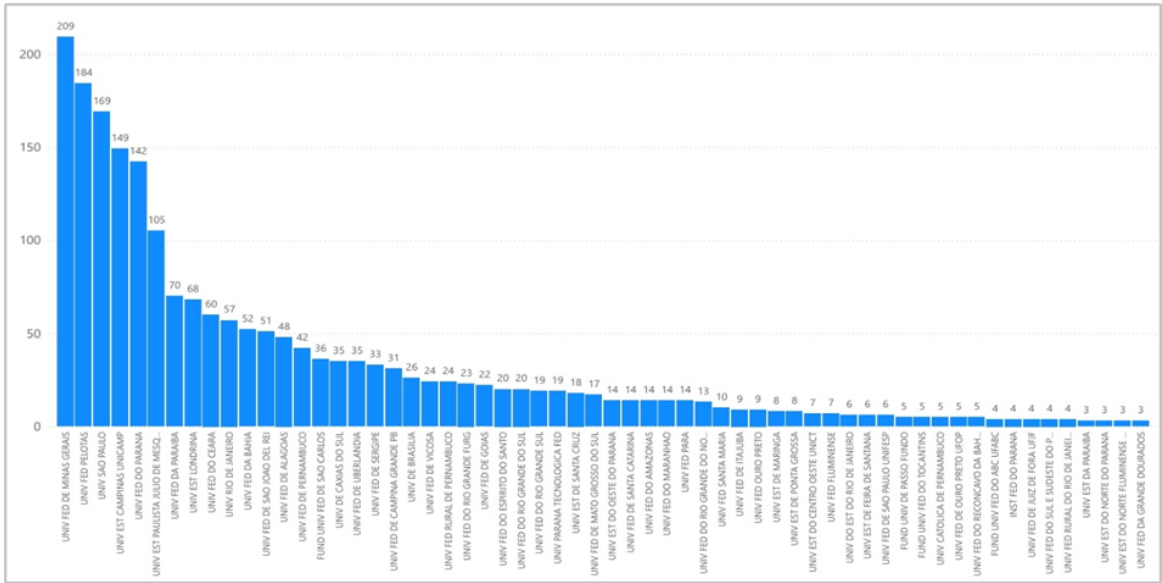
Gráfico 3 - Categorização dos Requerentes de Patentes



Fonte: Elaborado pelo autor (2025).

Os pedidos de patentes exclusivamente de universidades públicas somaram 2.003, conforme o Gráfico 4.

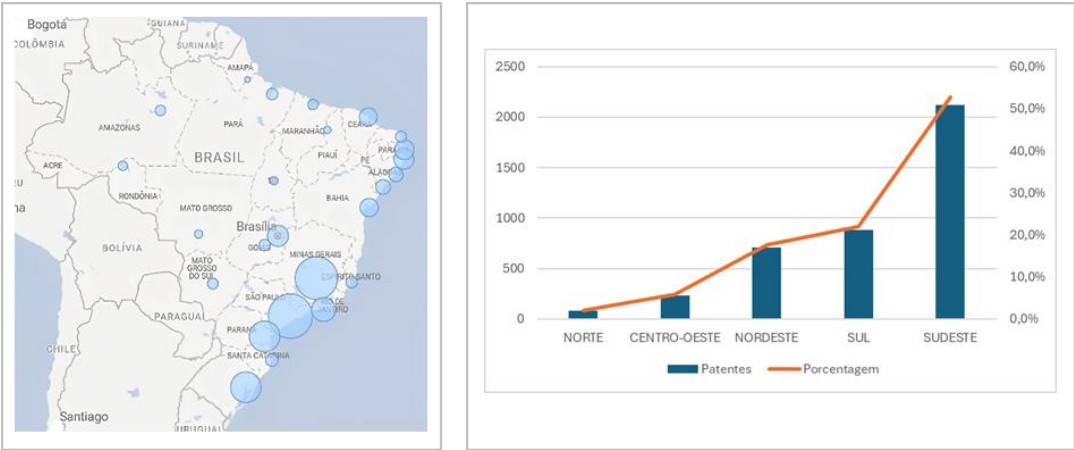
Gráfico 4 - Pedidos de Patentes de Universidades Públicas



Fonte: Elaborado pelo autor (2025).

A concentração dos pedidos de patentes representada pelos Gráficos 5a e 5b reflete-se na localização das universidades mais produtivas no período, com concentração nos estados do SUL e SUDESTE e em menor grau, no NORDESTE.

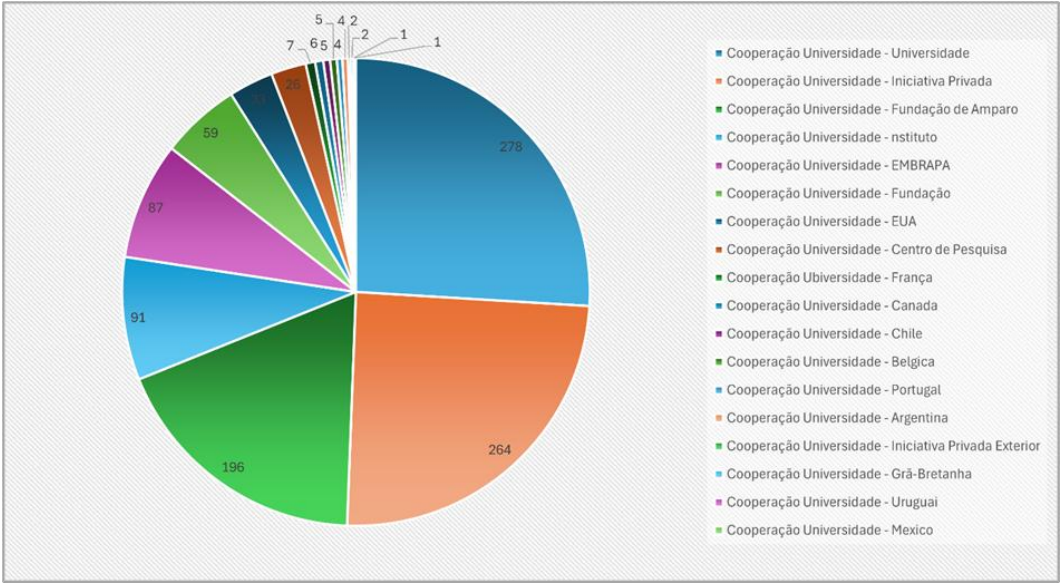
Gráfico 5a - Pedidos de Patentes por Estado Gráfico 5b - Pedidos de Patentes por Região



Fonte: Elaborado pelo autor (2025).

Outra constatação está relacionada às cooperações para pedidos de patentes, como pode ser observado no Gráfico 6.

Gráfico 6 - Pedidos de Patentes de Universidades em Cooperação



Fonte: Elaborado pelo autor (2025).

Percepções e Implicações

As percepções sobre o cenário das patentes de biotecnologia no Brasil entre 2014 e 2023 revelam flutuações significativas; embora a média de 401,9 pedidos de patentes por ano sugira um certo nível de atividade, a análise ano a ano mostra uma realidade mais instável. Em 2014, foram registrados 264 pedidos, com um aumento gradual até 2018, quando o número atingiu 602. No entanto, a partir de 2019, a tendência começou a oscilar, culminando em uma queda acentuada em 2023, com somente 162 pedidos. Essa variação indica que, apesar de alguns anos de crescimento, o setor enfrenta desafios que podem estar inibindo a inovação e a proteção de novas invenções.

A identificação de 29 patentes concedidas em 4.019, em um período de 10 anos (2014-2023), evidencia a morosidade do processo de concessão.

O avanço da biotecnologia no Brasil também é marcado pela colaboração entre diferentes entidades. Ao todo, 1.071 pedidos de patentes resultantes de colaborações foram registrados, mostrando a importância do trabalho em conjunto. As universidades, por exemplo, se destacam como fontes de inovação, tendo gerado 278 pedidos de patentes mediante parcerias. A interação de universidades com o setor privado também teve resultado frutífero, gerando 264 pedidos de patentes, mostrando como essa sinergia pode ajudar ao desenvolvimento de novas tecnologias.

Além disso, colaborações com fundações de amparo e institutos geraram 196 e 91 pedidos de patentes, respectivamente. A EMBRAPA, uma das principais instituições de pesquisa agrícola do Brasil, contribuiu com 87 pedidos de patentes em cooperação com universidades; essas iniciativas representam um esforço coletivo que pode transformar a paisagem da biotecnologia no país.

No cenário internacional, as colaborações com os Estados Unidos resultaram em 33 pedidos de patentes, enquanto parcerias com a França e o Canadá geraram 7 e 6 pedidos de patentes, respectivamente. Embora essas colaborações internacionais sejam menos numerosas, elas são essenciais para a troca de conhecimento e tecnologia, ampliando as possibilidades de inovação.

Entretanto, a dependência de inovações exteriores e a escassez de patentes na indústria farmacêutica são preocupações que emergem, sublinhando a necessidade de se afiar a capacidade de inovação do Brasil, no plano interior. A

complexidade, o moroso sistema de patentes, a burocracia e a falta de relação de dados escalonados, são barreiras que complicam o desenvolvimento do setor.

Esses dados não constituem meras estatísticas, mas refletem uma realidade concreta com implicações diretas para as políticas públicas e às práticas empresariais. Nesse sentido, é necessário criar um ambiente que, não somente proteja as invenções, como também estimule a pesquisa e o desenvolvimento. A análise das patentes oferece uma visão do estado da biotecnologia no Brasil e direciona as ações futuras destinadas a revitalizar o potencial inovador do país. A trajetória das patentes, neste setor, apresenta as dificuldades, solapando a militância em favor de uma necessidade que sinta urgência em um esforço coletivo que promova a inovação e que considere o futuro mais tranquilizante.

Os resultados obtidos neste trabalho permitiram a identificação de ações que podem contribuir significativamente para o fortalecimento da biotecnologia no Brasil, embora a maioria delas esteja além do escopo deste estudo e envolva aspectos relacionados a políticas, criação de plataformas, fundos de fomento e colaboração entre setores. Uma ação destacada é a documentação de casos de sucesso em patentes, destinada à avaliação das inovações.

Em segundo lugar, será igualmente importante um trabalho periódico relacionado às tendências em biotecnologia. Esses relatórios, a partir de dados de patentes, auxiliarão a mapear oportunidades de inovação.

Finalmente, não pode ser desprezada a promoção da pesquisa acadêmica, especialmente para estimular trabalhos que possam recorrer aos dados obtidos neste trabalho.

1.5. CONSIDERAÇÕES FINAIS E A TRANSIÇÃO NECESSÁRIA PARA UM MODELO PRESCRITIVO

O estudo revela que, apesar dos avanços na área de patentes de biotecnologia no Brasil entre 2014 e 2023, o país ainda enfrenta desafios significativos que limitam seu potencial inovador. A dependência de inovações estrangeiras, a burocracia do sistema de patentes e a falta de colaboração entre setores são barreiras que precisam ser superadas para o Brasil poder se destacar no cenário global de biotecnologia.

As consequências práticas das descobertas dessa pesquisa são voltadas para aqueles que formulam políticas, empresas e instituições de ensino, evidenciando ser

preciso um esforço coletivo para estabelecer um ambiente que não só resguarde as inovações, mas também incentive a pesquisa e o progresso.

Implementar políticas públicas que promovam a cooperação entre universidades e empresas, além de simplificar o processo de concessão de patentes, irá certamente revigorar o setor.

CAPÍTULO 2 - ARTIGO SOB TÍTULO: EFICIÊNCIA NO DESEMPENHO INOVADOR EM BIOTECNOLOGIA: UMA ANÁLISE ENVOLTÓRIA DE DADOS (DEA) NAS UNIVERSIDADES PÚBLICAS BRASILEIRAS⁶

O presente capítulo é baseado no artigo intitulado “Eficiência no desempenho inovador em biotecnologia: uma análise envoltória de dados (DEA) nas universidades públicas brasileiras”, mantendo integralmente os resultados e a estrutura do artigo publicado. Adicionalmente, incorpora uma seção complementar (seção 2.6) que explora um subproduto do modelo DEA (cujos fundamentos teóricos foram apresentados na Seção 2.7 do REFERENCIAL TEÓRICO (CATEGORIAS PRINCIPAIS) p. 35): a matriz de lambdas (λ), a matriz de folgas (slacks) e a DMU Virtual. Este aprofundamento, viabilizado pelo acesso à matriz completa de pesos do modelo, permite identificar os *benchmarks* factíveis específicos para a UFAM e quantificar as folgas de *outputs*, informações que, embora derivadas do mesmo modelo, não faziam parte do escopo original da publicação, mas são relevantes para o diagnóstico institucional proposto nesta tese. Assim, a seção 2.6 deve ser lida como um desdobramento analítico da tese, preservando a integridade do artigo publicado como unidade autônoma.

2.1. INTRODUÇÃO: EFICIÊNCIA INOVADORA E O PAPEL DA DEA NAS UNIVERSIDADES PÚBLICAS BRASILEIRAS

É a transferência de tecnologia que catalisa o progresso, convertendo inovações científicas em benefícios socioeconômicos. A habilidade de transformar inovação em competitividade e, ao mesmo tempo, em ganhos para a sociedade em um mundo cada vez mais interconectado é um fator-chave para o desenvolvimento sustentável, particularmente em áreas estratégicas e complexas como a biotecnologia.

Quando universidades e empresas se unem, criam um ecossistema que vai além da inovação tecnológica e que fornece soluções reais para os problemas sociais

⁶ O presente capítulo é uma versão reformatada do artigo intitulado “Eficiência no desempenho inovador em biotecnologia: uma análise envoltória de dados (DEA) nas universidades públicas brasileiras”, publicado na Revista Peer Review, Vol. 8, 4. Y, p. 205–223, 2026, DOI: 10.53660/PRW-2922-5210, (KUME; CHAVES; OLIVEIRA JÚNIOR, 2026). O texto base das seções 2.1 a 2.5 é reproduzido integralmente, conforme normas de autocitação e tese-artigo, com adaptações estruturais para adequação ao formato da tese.

e econômicos do nosso tempo. A Organização para a Cooperação e Desenvolvimento Econômico (OCDE) ressalta que, sem essa ligação, o potencial inovador não é totalmente realizado, restringindo o crescimento econômico e a qualidade de vida. Em um país como o Brasil, que possui vasto potencial biotecnológico, entender essa dinâmica vai além de uma opção acadêmica: trata-se de uma urgência estratégica para direcionar políticas públicas e reforçar os ecossistemas de inovação.

Para os fins deste trabalho, entendemos por transferência de tecnologia o trânsito do conhecimento dos laboratórios para o mercado, onde ele se converte em produtos e serviços capazes de gerar valor científico, econômico e social. A inovação é, portanto, a realização: a implementação real de ideias que geram resultados concretos. Surge, então, a noção de eficiência acadêmica, que estuda a transformação que as instituições de ensino superior promovem ao converter recursos e pesquisas em patentes, tecnologias e ativos de conhecimento em benefício da sociedade.

É, portanto, necessário avaliar em que medida as atividades de pesquisa das instituições de ensino superior se refletem na geração de propriedade intelectual (PI), requerendo o uso de métodos capazes de relacionar os insumos disponíveis aos resultados observados. Assim sendo, o objetivo não é afirmar uma relação causal entre pesquisa e propriedade intelectual, mas reconhecer que esses elementos podem, potencialmente, estabelecer a posição relativa das instituições na fronteira de eficiência vinculada à inovação e à transferência tecnológica.

Ao apontar as instituições que se destacam em eficiência, a pesquisa oferece um *benchmark* que pode ser utilizado na formulação de estratégias e políticas públicas que ampliem a relevância das iniciativas de P&D em biotecnologia e que fortaleçam a conexão entre produção acadêmica e desenvolvimento econômico.

A escolha de restringir a amostra a instituições públicas se fundamenta na busca pela homogeneidade do modelo, de modo que a comparação seja feita entre entidades que atuam sob os mesmos regimes jurídicos, que lidam com desafios estruturais semelhantes e que têm condições de financiamento e funcionamento comparáveis.

A presente investigação delimita o seu escopo central na mensuração da eficiência das Instituições de Ensino Superior (IES) quanto à gênese de ativos de inovação. Almeja-se que este diagnóstico transcenda a contribuição teórica ao estado da arte, consolidando-se como um referencial estratégico para a gestão universitária.

2.2. REFERENCIAL TEÓRICO: INOVAÇÃO, TRANSFERÊNCIA TECNOLÓGICA E DEA

Inovação e transferência tecnológica

A dinâmica de desenvolvimento em setores de fronteira tecnológica, com ênfase na biotecnologia, encontra na transferência de tecnologia e na inovação os seus eixos de sustentação. É nesse cenário que a colaboração entre IES e o setor produtivo ultrapassa o intercâmbio passivo de informações; trata-se de um mecanismo determinante para converter conhecimento científico em ativos de mercado e otimizar trajetórias de pesquisa. Segundo O'Dwyer, Filieri e O'Malley (2023), essa interação é tão mais eficaz quando depende não só da complementaridade de conhecimentos, mas do aporte estratégico de recursos, variáveis que balizam a competitividade e a maturação industrial em cadeias produtivas correlatas, como a farmacêutica.

No campo biotecnológico, essa integração atua como o vetor principal para gerar produtos de elevado valor agregado. O desenvolvimento tecnológico na biotecnologia não é apenas uma questão de funcionalidade entre inovação e transferência, mas sim da administração das tensões que surgem nos regimes de padronização. Segundo Langer (2023), os padrões funcionam como ferramentas de governança que influenciam o comportamento das empresas por meio de uma interação dialética entre o empoderamento e a limitação. Sob esta ótica, as normas institucionais não são parâmetros neutros; elas constituem mecanismos de poder que conferem às organizações a adaptabilidade necessária para mitigar pressões externas, ao mesmo tempo, em que estabilizam arranjos cooperativos voltados à atividade inventiva. Assim, a consolidação de ecossistemas de inovação pressupõe a capacidade institucional de converter diretrizes normativas em recursos estratégicos para a geração de ativos de elevado valor agregado.

Apesar desse potencial, a conversão de descobertas laboratoriais em aplicações práticas no campo da biotecnologia ainda encontra barreiras significativas. As diferenças culturais e as lacunas na gestão da propriedade intelectual são os principais fatores que geram resistência a esse fluxo de conhecimento. É necessário um modelo de governança que alinhe a produção acadêmica às necessidades de escalonamento do mercado para superar tais entraves (Awasthy *et al.*, 2020).

No contexto dos Núcleos de Inovação Tecnológica (NIT), os indicadores assumem relevância, sobretudo aqueles relacionados à gestão da propriedade intelectual e às atividades de transferência de tecnologia. O número de depósitos de patentes é frequentemente utilizado como parâmetro de desempenho, mas, tomado isoladamente, não representa a complexidade dos processos de inovação nas instituições. A interpretação desse dado exige articulação com outros elementos que influenciam a atuação dos NITs, como a qualificação das equipes, a presença de profissionais especializados e o suporte técnico terceirizado. Esses aspectos afetam diretamente a capacidade de transformar resultados de pesquisa em tecnologias transferíveis e ajudam a compreender por que algumas instituições avançam mais rapidamente na consolidação de uma cultura de inovação (Lima *et al.*, 2023).

Análise Envoltória de Dados (DEA)

A DEA é uma metodologia sólida para avaliar a eficiência de unidades, servindo para diversos setores, entre eles produtivos e serviços. O método tem uma abordagem de programação não linear para múltiplos insumos e produtos, oferecendo uma medida quantitativa de eficiência, sendo aplicável em diferentes campos, revelando seu potencial para mensurar o desempenho em condições diversas, incluindo transferências tecnológicas e desenvolvimento de novos processos (Charnes; Cooper; Rhodes, 1978).

No que diz respeito às IES, a DEA é aplicada na avaliação da eficiência do ensino, possibilitando comparações entre docentes com base em indicadores de entrada e saída, tais como formação acadêmica, tempo de serviço e avaliações de desempenho. A eficiência pedagógica é o que se busca aqui, e não uma aplicação direta em biotecnologia ou em transferência de tecnologia, mas o método se evidencia apto a lidar com diversas configurações de entradas e saídas (Lu; Yan Bai, 2012; Nazarko; Šaparauskas, 2014).

Representações bidimensionais da DEA, compatíveis com os modelos CCR e BCC, facilitam a visualização da posição das unidades em relação à fronteira eficiente e auxiliam na interpretação comparativa do desempenho institucional. Embora a DEA opere, em sua formulação completa, com múltiplas variáveis e uma fronteira que assume a forma de uma superfície multidimensional, projeções em duas dimensões

permitem comunicar de maneira mais acessível como cada unidade se distancia ou se aproxima dessa fronteira teórica (Costa; Mello; Meza, 2016).

Além disso, os modelos DEA com retornos variáveis de escala (VRS/BCC) possibilitam comparar instituições de tamanhos distintos, evitando que diferenças de porte entre as instituições influenciem os resultados da análise. Nesses modelos, a eficiência é mensurada via relação insumos contra produtos (e não pelo volume absoluto dos insumos), evitando assim o favorecimento de organizações maiores. Revisões sistemáticas da aplicação da DEA no ensino superior confirmam esta robustez da metodologia (Ahmad; Khurizan, 2024), assim como outros estudos, que ressaltam que o VRS é apropriado para situações institucionalmente heterogêneas (Emrouznejad; Yang, 2018).

Indicadores de desempenho em IES

Indicadores de desempenho são essenciais para compreender como as universidades convertem insumos em resultados no campo da inovação. Abordagens que integram indicadores combinados com técnicas de análise de eficiência, como a DEA, são o caminho para realizar essa tarefa, permitindo que sejam observados, simultaneamente, aspectos do ensino, da pesquisa e da transferência de tecnologia, tornando o diagnóstico institucional mais preciso. Entre essas metodologias, é possível adotar diferentes formas de interpretar os resultados. Em alguns casos, trabalha-se com cenários compensatórios, nos quais um desempenho favorável em determinada dimensão pode suavizar limitações observadas em outra. Em outros casos, escolhem-se cenários não compensatórios, em que é necessário um equilíbrio entre todos os indicadores, de modo que um resultado alto não esconda fragilidades particulares (El Gibari; Gómez; Ruiz, 2022).

A aplicação da DEA em universidades evidencia que a seleção dos indicadores vai além do uso administrativo cotidiano e exige a identificação de variáveis que expressem, de modo mensurável, como os insumos institucionais são convertidos em resultados acadêmicos, científicos ou tecnológicos, sendo que a seleção criteriosa das variáveis permitirá que o modelo identifique discrepâncias de desempenho entre instituições que operam em contextos semelhantes (Pinheiro *et al.*, 2022).

Análise de Eficiência

Com os indicadores definidos, a literatura passa a explorar como esses elementos podem ser avaliados integradamente por meio de modelos de eficiência.

Uma avaliação de eficiência com um modelo dinâmico de DEA, aplicado a 17 centros de pesquisa da Academia Chinesa de Ciências entre 2012 e 2015, demonstrou que esse método obteve melhores resultados do que os modelos estáticos, em termos de avaliação de eficiência em P&D. Os dados indicaram um posterior crescimento na transferência de tecnologia e na inovação científica, porém ainda há espaço para aumentar o impacto dos resultados da pesquisa (Xiong; Yang; Guan, 2018).

Pesquisas similares com o emprego da DEA, também na China, reforçam a necessidade de otimizar processos internos e validar a relação entre eficiência em pesquisa e inovação, destacando a transferência de tecnologia (Liang; Yin, 2024).

As pesquisas brasileiras que aplicam a DEA em universidades tendem a concentrar-se em elementos tradicionais de desempenho acadêmico, como o nível de qualificação do corpo docente, o investimento financeiro por estudante, os índices de conclusão nos cursos de graduação e as avaliações da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Tais análises conseguem destrinchar práticas de eficiência e expor as vulnerabilidades das instituições no âmbito do ensino e da pós-graduação. Embora esses estudos não tenham abordado temas como inovação, propriedade intelectual ou biotecnologia, eles ressaltam como a DEA é fundamental para elucidar como as IES transformam insumos em resultados que impulsionam o desenvolvimento do tecido socioeconômico (Cohen; Paixão; Oliveira, 2018; Pereira, 2020a).

A discussão sobre eficiência nas instituições de ensino superior é essencial para compreender como os recursos públicos são transformados em resultados acadêmicos e sociais. Um estudo que avaliou universidades federais brasileiras com base em DEA incorporou dimensões de ensino, pesquisa, extensão e inovação, esta última representada pelo registro de patentes. Os resultados evidenciaram diferenças regionais expressivas, mas também indicaram uma tendência de melhora ao longo do tempo, sugerindo que políticas adaptadas às realidades locais são essenciais (Letti; Bittencourt; Vila, 2020).

No campo da inovação regional, estudos que aplicaram DEA para medir a eficiência dos sistemas de inovação mostraram que fatores como o volume de produtos inovadores e o investimento em P&D são determinantes para o desempenho. Mesmo em regiões que enfrentaram cortes de recursos, houve crescimento nos indicadores de inovação, reforçando a importância de políticas públicas voltadas à melhoria contínua e ao uso criterioso dos recursos disponíveis (Firsova; Chernyshova, 2020).

Ferramentas

O R continua sendo uma linguagem de programação voltada à análise estatística, modelagem e visualização de dados, criada em 1993 no Departamento de Estatística da Universidade de Auckland e mantida até hoje por uma comunidade internacional que desenvolve tanto o núcleo da linguagem quanto seus pacotes. Esses pacotes reúnem funções, conjuntos de dados e código compilado que ampliam as capacidades do ambiente e permitem sua aplicação em áreas como ciência de dados, bioinformática, econometria e aprendizado de máquina (Grolemund, 2018). O RStudio, por sua vez, consolidou-se como o principal ambiente integrado de desenvolvimento para trabalhar com R, oferecendo recursos de edição, execução, visualização e documentação em uma interface unificada. A mudança mais recente ocorreu na esfera institucional: a empresa responsável pelo RStudio passou a se chamar Posit (POSIT PBC, 2022), refletindo a ampliação do seu escopo para além do R e incorporando também ferramentas voltadas ao uso de Python, ao desenvolvimento com Quarto e a outras tecnologias de ciência de dados. Essa alteração não modifica o funcionamento do RStudio IDE, que continua disponível e amplamente utilizado, mas reposiciona a empresa em um ecossistema mais amplo de *software* científico e de código aberto.

Quadro 1 - Bibliotecas R a serem utilizadas

Biblioteca	Função
<i>deaR</i>	Conjunto de funções para Análise Envoltória de Dados
<i>ggplot2</i>	Um sistema para criar gráficos descritivamente
<i>dplyr</i>	Ferramenta para trabalhar com estrutura de dados bidimensional
<i>readxl</i>	Obtém dados de arquivos .xlsx

Fonte: Elaborado pelo autor (2026).

2.3. METODOLOGIA: MODELO DEA EM DOIS ESTÁGIOS METODOLÓGICOS

A DEA é uma metodologia não paramétrica baseada em programação linear, utilizada para avaliar a eficiência relativa de Unidades Tomadoras de Decisão (DMUs) que operam com múltiplos insumos (*inputs*) e produtos (*outputs*). No contexto deste estudo, o objetivo da DEA é mensurar como as universidades brasileiras otimizam seus recursos para a geração de resultados em biotecnologia e transferência de tecnologia.

Diferente de abordagens estatísticas tradicionais que focam no comportamento médio, o DEA constrói uma "fronteira de eficiência" baseada nas melhores práticas observadas na amostra. As instituições situadas nesta fronteira recebem o escore 1.0 (100% de eficiência), servindo de *benchmark* para as demais.

Considerando a heterogeneidade estrutural das universidades públicas brasileiras, que variam amplamente em porte, capacidade instalada e volume de recursos, adotou-se o modelo com retornos variáveis de escala (VRS/BCC). Assim, o modelo identifica a eficiência relativa de cada DMU dentro de sua própria realidade operacional, proporcionalmente ao uso dos recursos, assegurando que diferenças de escala não interfiram na comparação entre as instituições, o que é particularmente adequado para estudos envolvendo sistemas universitários diversos.

Para conferir maior rigor e poder de discriminação à análise, esta pesquisa adotou uma modelagem em dois estágios sequenciais, executada em ambiente R (pacote *deaR*):

- Primeiro Estágio (Eficiência Técnica): aplicou-se o modelo DEA-VRS (Retornos Variáveis de Escala) com orientação ao *output*. O uso do modelo VRS de Banker, Charnes e Cooper (1984) justifica-se pela heterogeneidade das IES brasileiras, garantindo que instituições de diferentes portes sejam comparadas dentro de suas respectivas realidades de escala. Nesta fase, os *inputs* foram: Graduados em Biotecnologia, Bolsas e Projetos CAPES; e os *outputs*: Produção Científica (SCOPUS E PUBMED) e Depósitos de Patentes.
- Segundo Estágio (Super-Eficiência e Desempate Qualitativo): as instituições que atingiram a fronteira de eficiência no primeiro estágio (escore $\geq 0,9999$) foram submetidas ao modelo de Super-Eficiência de (Andersen; Petersen, 1993). Neste estágio, introduziu-se o indicador RUF-Inovação (*Ranking* Universitário Folha)

como uma variável de saída qualitativa adicional (RUF, 2025). Este procedimento permitiu o desempate (*tie-breaking*) das unidades de elite e a identificação do *benchmark* absoluto do setor, integrando o volume produtivo ao impacto inovador institucional.

Essa abordagem permite não apenas identificar a eficiência relativa, mas também isolar o efeito da escala, funcionando como um guia preciso para potencializar as políticas de inovação nas IES (Charnes; Cooper; Rhodes, 1978). Para garantir a comparabilidade visual e mitigar o efeito de outliers de super-eficiência, os resultados foram submetidos a uma normalização logarítmica, assegurando que a escala de análise reflita fielmente o distanciamento entre a fronteira de excelência e as demais unidades.

Bases de Dados

As bases de dados reúnem informações provenientes de fontes reconhecidas, cada uma essencial para a construção e análise do estudo no contexto das universidades públicas brasileiras.

Graduados em Biotecnologia

Utilizaram-se dados do Censo da Educação Superior, o qual é realizado todos os anos pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), abrangendo o período de 2014 a 2023 (INEP, 2020), que servirão como input para a análise.

Publicações (SCOPUS e PUBMED)

Publicações científicas nas bases de dados SCOPUS e PUBMED, amplamente reconhecidas e relevantes no campo da biotecnologia, entre 2014 e 2023, oriundas de pesquisadores de universidades públicas brasileiras. Pesquisou-se utilizando termos bem definidos que se relacionam ao tema do estudo, o que assegurou a relevância e a qualidade do material examinado (PUBMED, 2019; SCOPUS, 2019).

Fomento e Projetos CAPES

Com base na Política de Dados Abertos (Decreto n.º 8.777/2016), foram extraídos dados da CAPES (2013-2022) referentes a:

- Bolsas: concessões das modalidades de pós-graduação vinculadas à Diretoria de Programas e Bolsas no País (DPB);
- Projetos: informações sobre os Projetos dos Programas de Pós-Graduação *stricto sensu*, incluindo área de concentração e indicativo de fomento.

Categoria de Inovação do Ranking Universitário Folha (RUF)

A introdução do componente de Inovação do RUF como variável de saída (*output*) no segundo estágio da análise (Super-DEA) justifica-se pela necessidade de um critério de desempate que considere a infraestrutura tecnológica e o impacto. A escolha metodológica tem relevância nesta perspectiva pelos seguintes motivos:

- a. Procedência e Auditabilidade dos Dados: embora publicado por um veículo de comunicação, o indicador de Inovação do RUF não é subjetivo. Ele é construído a partir de dados brutos e auditáveis provenientes do Instituto Nacional da Propriedade Industrial (INPI) e de bases de produção científica internacional (como a *Web of Science*). Ele contabiliza não apenas o volume de depósitos de patentes, mas também o percentual de patentes em parceria com o setor produtivo, servindo como um reflexo fiel direto para a Transferência de Tecnologia (TT);
- b. Validação por Órgãos de Controle e Gestão: a legitimidade do indicador é ratificada por sua adoção em Planos de Desenvolvimento Institucional (PDI) e Relatórios de Gestão de universidades de excelência, como Universidade de São Paulo (USP), Universidade de Brasília (UnB), Universidade Federal do Rio de Janeiro (UFRJ) e Universidade Federal do Rio Grande do Sul (UFRGS). Tais instituições utilizam esses critérios para balizar metas de desempenho, elevando o status da métrica de um ranking informativo para um indicador de gestão pública reconhecido;
- c. Certificação Internacional (IREG): o RUF é membro do IREG *Observatory on Academic Ranking and Excellence*. Isso significa que sua metodologia segue

diretrizes internacionais de transparência e rigor, as mesmas aplicadas a rankings globais como o QS e o *Times Higher Education* (THE);

- d. Função de Diferenciação Qualitativa: no modelo DEA proposto, o primeiro estágio mede a eficiência produtiva quantitativa. O RUF-Inovação entra no segundo estágio para premiar a eficiência de impacto. Ele permite distinguir instituições que possuem um ecossistema de inovação mais maduro (NITs estruturados, cultura de parceria universidade-empresa), funcionando como um filtro de qualidade indispensável para o escalonamento das unidades situadas na fronteira de eficiência."

Patentes

A base de dados de pedidos de patentes utilizada para a pesquisa foi o Latipat⁷, de 2014 a 2023 para este estudo, um ponto de acesso público e gratuito, permitindo a consulta de informações técnicas em espanhol e português em documentos de patentes da América Latina e Espanha. A pesquisa foi direcionada para pedidos de patentes na área de biotecnologia, utilizando o conjunto de Classificações Internacionais de Patentes (CIP) de biotecnologia, definido como Biotec BR, por pesquisadores do INPI (Weid *et al.*, 2018). A análise incluiu documentos associados a pesquisadores de universidades públicas brasileiras, garantindo a relevância do material para o contexto do estudo.

Quadro 2 - Resumo das Variáveis e Modelagem

Variáveis (2014-2023)	Origem	Classificação	Variável	Estágio DEA	Justificativa
Graduações em Biotecnologia	INEP	Educação	<i>Input</i>	1 e 2	Capacidade de formação de RH.
Publicações (SCOPUS/PUBMED)	SCOPUS PUBMED	Produção	<i>Output</i>	1 e 2	Base científico para PI.
Bolsas de Pós-Graduação	CAPES	Educação	<i>Input</i>	1 e 2	Fomento à formação de pesquisadores.
Projetos de Pós-Graduação	CAPES	Pesquisa	<i>Input</i>	1 e 2	Infraestrutura para inovação.
Patentes de Biotecnologia	INPI Patentscope	Prop. Intelectual	<i>Output</i>	1 e 2	Potencial de TT.
Índice de Inovação (RUF)	Folha/RUF	Ranking	<i>Output</i>	Apenas 2	Desempenho institucional.

Fonte: Elaborado pelo autor (2026).

⁷ LATIPAT. Latipat - Espacenet. [S. l.], 2017. Disponível em: <https://lp.espacenet.com>. Acesso em: 21 mar. 2025.

Etapas da análise

Para assegurar a homogeneidade, aplicou-se um protocolo de seleção sobre o universo inicial de 71 IES. Foram excluídas instituições sem programas de pós-graduação ativos na área ou com dados inconsistentes. A amostra final consolidou-se em 29 DMUs, garantindo estabilidade estatística frente ao número de variáveis.

Etapa 1 - DEA com Retornos Variáveis de Escala (VRS)

Focada na maximização de *outputs* com base nos recursos disponíveis, esta fase respeita as diferenças de investimento e porte das IES brasileiras. As instituições eficientes formam a primeira fronteira de comparação.

Etapa 2 - Super-DEA com Retornos Constantes de Escala (CRS)

Nesta fase, aplica-se o modelo de Andersen e Petersen (1993) às unidades que atingiram o escore máximo. A introdução do RUF-Inovação permite premiar instituições com ecossistemas de inovação mais maduros e maior interação universidade-empresa. O uso de Retornos Constantes de Escala (CRS) nesta fase visa identificar a eficiência de escala ótima, isolando os *benchmarks* do setor.

Normalização dos Dados e Representação Visual

Os valores de eficiência foram submetidos a uma normalização logarítmica, utilizando a fórmula:

$$Score_{norm} = \frac{\ln(Score + 1)}{\ln(Max + 1)}$$

Esta técnica foi fundamental para preservar a sensibilidade analítica das instituições de menor porte frente ao desempenho exponencial da eficiente (USP), garantindo que as cinco faixas de eficiência reflitam a realidade competitiva do setor sem distorções de escala:

- Baixa Eficiência (0 a 0,24) - Desempenho relativo mínimo.
- Moderada Eficiência (0,25 a 0,49) - Demanda esforços significativos.

- Média Alta Eficiência (0,50 a 0,74) - Progresso, mas ainda exige atenção.
- Alta Eficiência (0,75 a 0,99) - Solidez, mas requer otimização.
- Eficiente (1,00) - O desempenho ideal.

Para a representação gráfica dos dados, foram associadas a um gradiente de cores de vermelho a verde, facilitando a comunicação visual dos resultados.

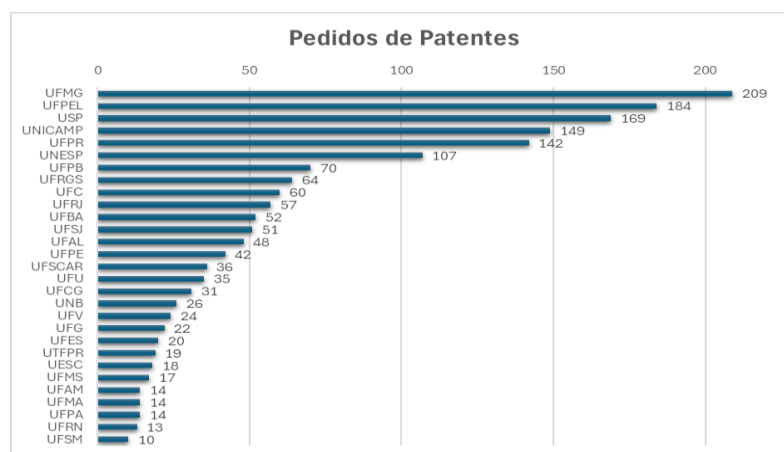
Entre as IES brasileiras, as minúcias da eficiência foram obtidas com o auxílio do modelo DEA, em duas etapas. A avaliação inicial com retornos variáveis de escala (VRS) foi detalhada, e a inclusão do Super-DEA, juntamente com o RUF-Inovação, adicionou uma qualidade essencial à análise. As faixas de eficiência definidas categoricamente fortalecem a intuição sobre os dados e ajudam a localizar *benchmarks* para transferência de tecnologia.

2.4. RESULTADOS: EFICIÊNCIA INOVADORA DAS UNIVERSIDADES PÚBLICAS ANALISADAS

A pesquisa dos pedidos de patente de biotecnologia por requerentes brasileiros, para o período de 2014 a 2023, apresenta 2.003 ocorrências exclusivamente para universidades públicas.

Desses 2.003 pedidos de patentes, foram selecionadas as 29 IES com maior número de pedidos no período analisado, conforme a Figura 1.

Figura 1 - Pedidos de Patentes de Biotecnologia - IES Públicas



Fonte: Elaborado pelo autor (2026).

A análise DEA realizada com os dados do Quadro 3 permitiu identificar as instituições que operam na fronteira eficiente, bem como aquelas que apresentam ineficiência relativa em relação ao uso dos recursos disponíveis.

Quadro 3 - Tabela de dados para DEA

IES	Graduados em Biotecnologia	Publicações SCOPUS	Publicações PUBMED	Bolsas CAPES	Projetos CAPES	RUF Inovação	Patentes
UFMG	633	144	18	4988	1305	3,65	209
UFPEL	212	37	4	1457	1383	2,60	184
USP	67	614	125	8199	985	3,95	169
UNICAMP	445	239	55	2800	139	3,62	149
UFPR	124	130	10	4054	573	3,70	142
UNESP	283	272	177	3017	809	3,21	105
UFPB	213	41	8	346	45	2,15	70
UEL	1	58	2	1630	217	2,66	68
UFC	240	94	19	2418	819	2,74	60
UFRJ	348	221	16	5818	1423	3,87	57
UFBA	439	52	12	877	916	2,75	52
UFSJ	39	30	2	217	183	2,31	51
UFAL	67	32	2	356	437	2,17	48
UFPE	223	101	13	2224	412	2,80	38
UFSCAR	460	94	14	1536	1199	2,83	36
UFU	382	37	6	1197	221	3,35	35
UFCG	120	8	1	87	516	2,44	31
UNB	379	130	32	1875	1339	3,05	26
UFV	266	130	11	633	611	3,80	24
FURG	141	83	24	619	313	1,71	23
UFG	207	44	6	1285	667	2,86	22
UFES	691	34	5	1078	337	3,15	20
UFRGS	140	95	10	3095	505	3,58	20
UTFPR	207	54	4	63	179	2,94	19
UESC	146	27	1	322	409	1,58	18
UFMS	256	29	2	170	115	2,54	17
UFAM	311	30	3	739	817	1,86	14
UFMA	141	21	4	157	371	2,15	14
UFPA	188	56	14	2146	529	3,68	14
UFRN	301	52	10	2091	367	3,08	13
UFSM	194	91	14	1445	46	2,93	10

Fonte: Elaborado pelo autor (2026).

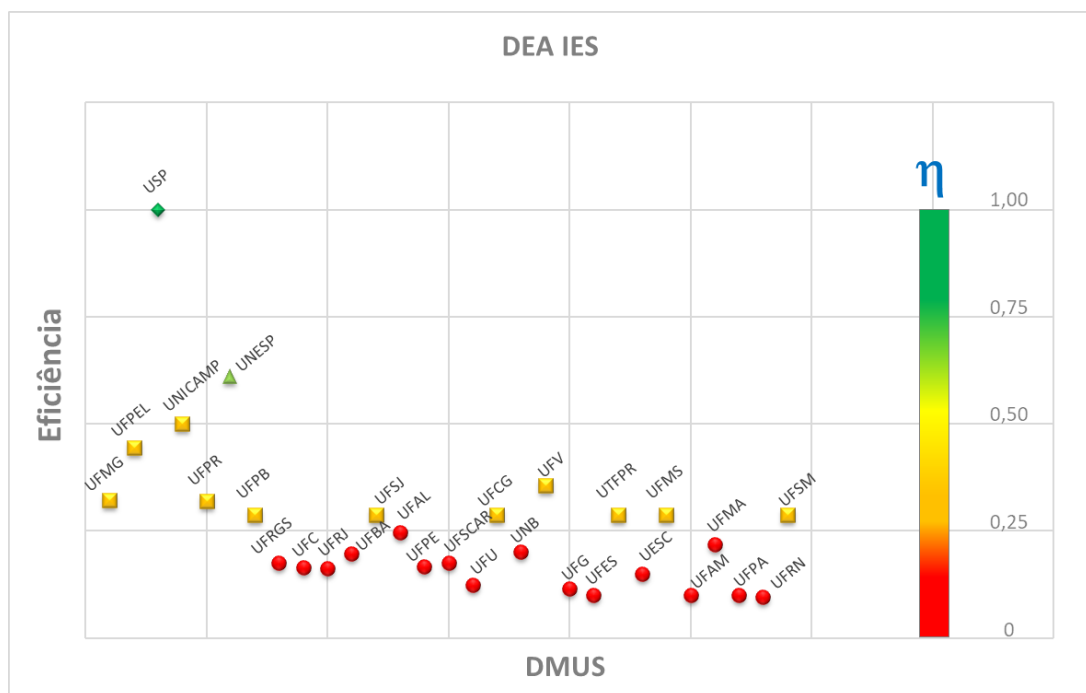
Os resultados fornecem subsídios para orientar decisões estratégicas na promoção da transferência tecnológica e na definição de políticas para fortalecer a inovação em biotecnologia.

O gráfico da Figura 2 apresenta a eficiência das IES distribuída em cinco faixas de desempenho. Os resultados revelam que o volume absoluto de depósitos não é garantia de eficiência máxima. Embora a Universidade Federal de Minas Gerais (UFMG) lidere em termos quantitativos de patentes, ela se posiciona na faixa de Alta

Eficiência (0,7966), abaixo da fronteira unitária. Isso ocorre porque o modelo DEA-VRS identifica como ineficiência relativa à subutilização proporcional de insumos: a UFMG detém um volume massivo de Bolsas e Projetos CAPES (Quadro 3) que, se comparado ao desempenho da USP (1,0000), indica que a sua taxa de conversão por unidade de recurso ainda possui margem para otimização, especialmente na geração de ativos de alto impacto e interação com o mercado (mensurada pelo RUF-Inovação).

No contexto desta base, a USP consolida-se como a única unidade plenamente eficiente (*benchmark*) por apresentar o melhor equilíbrio entre a produção científica e a proteção de ativos, catalisada por um ecossistema de inovação superior. É importante destacar que a metodologia Super-DEA agiu como um filtro de qualidade: ao integrar o indicador RUF-Inovação, o modelo diferencia a 'produção por volume' da 'produção por impacto'. Assim, instituições que possuem grandes estruturas, mas menor inserção no setor produtivo, têm seus escores ajustados, revelando que a eficiência técnica é independente da soberania orçamentária.

Figura 2 - Eficiência na Transferência Tecnológica



Fonte: Elaborado pelo autor (2026).

Os resultados foram submetidos a uma harmonização logarítmica, transformação estatística aplicada para estabilizar a variância decorrente da heterogeneidade entre as IES e permitir uma visualização clara das faixas de

eficiência. Ressalta-se que a linearização logarítmica preserva a ordenação original dos escores, garantindo a integridade dos resultados ao mesmo tempo que evidencia o distanciamento real entre a fronteira de eficiência e as demais unidades. Este procedimento, aliado à alternância entre os modelos VRS e CRS nas duas etapas do DEA, torna os resultados comparáveis sem introduzir distorções por diferenças de porte das DMUs. A distinção entre eficiência técnica e eficiência de escala permite uma leitura precisa: enquanto a eficiência técnica indica o desempenho da instituição na otimização dos recursos disponíveis, a eficiência de escala revela se o porte da instituição é adequado ao seu nível de operação. Assim, o modelo assegura que a liderança da USP e de outras IES de grande porte não seja um mero reflexo de sua escala, mas sim de uma eficiência técnica superior na conversão de insumos em inovação.

Os resultados do modelo DEA se mostram relevantes e balizadores, devendo ser complementados por outras métricas que auxiliem no alavancamento da inovação com transferência tecnológica.

2.5. CONCLUSÃO: SÍNTESE DA EFICIÊNCIA INOVADORA NAS IES BRASILEIRAS

A avaliação da eficiência das instituições públicas brasileiras que atuam em biotecnologia e transferência de tecnologia revela um sistema heterogêneo, no qual poucos centros concentram maturidade inovadora enquanto a maioria das universidades opera abaixo do seu potencial.

A análise DEA em dois estágios, mostra que a eficiência inovadora não decorre exclusivamente do volume de recursos ou da produção científica, mas da capacidade institucional de transformar conhecimento em ativos protegidos e de possibilitar o estabelecimento de conexões efetivas com o setor produtivo.

O panorama identificado sugere que a criação de um ambiente favorável à inovação é um resultado que depende de políticas públicas destinadas a reduzir as disparidades regionais, na medida que mostra que os entraves são mais complexos do que a mera falta de financiamento, revelando uma natureza estrutural que, entre outros aspectos, contempla assimetrias entre instituições.

O estudo contribui ao indicar que a mensuração da eficiência em biotecnologia exige abordagens que combinem desempenho científico, capacidade de proteção

intelectual e impacto institucional. Futuras pesquisas podem aprofundar essa análise ao incorporar outros métodos de ciência de dados, indicadores de licenciamento, métricas de impacto econômico e modelos dinâmicos que capturem a evolução temporal da eficiência, ampliando a compreensão sobre os fatores que impulsionam a inovação no ambiente universitário brasileiro.

2.6. ANÁLISE COMPLEMENTAR: MATRIZ DE LAMBDA E DMU VIRTUAL DA UFAM

Os fundamentos teóricos da matriz de lambdas, da matriz de folgas e da DMU Virtual foram apresentados na Seção 2.7 p. 35. Nesta etapa, esses elementos são aplicados ao modelo DEA-VRS orientado a *outputs* desenvolvido no artigo, com foco na posição relativa da UFAM entre as 29 instituições analisadas.

Extração da matriz de lambdas (λ)

A matriz de lambdas foi extraída a partir do modelo DEA utilizando a função `lambdas()` do pacote `deaR` (Coll-Serrano; Bolós; Suárez, 2022). A matriz completa (29 × 29) encontra-se no ANEXO 1 - B, acompanhada da matriz de folgas (ANEXO 1 - C) e dos escores de eficiência (ANEXO 1 - A).

A linha correspondente à UFAM revela três *benchmarks* factíveis ($\lambda > 0,001$), apresentados no Quadro 4.

Quadro 4 - Matriz de lambdas da UFAM ($\lambda > 0,001$)

IES <i>benchmark</i>	Peso (λ)	Percentual na DMU Virtual
UFV (Universidade Federal de Viçosa)	0,56674	56,7%
UFPB (Universidade Federal da Paraíba)	0,33939	33,9%
UNICAMP (Universidade Estadual de Campinas)	0,09387	9,4%
Soma	1,00000	100%

Fonte: Elaborado pelo autor (2026).

As demais 26 instituições apresentaram $\lambda = 0$. Isso indica que apenas UFV, UFPB e UNICAMP possuem combinações de *inputs* e *outputs* compatíveis com a estrutura produtiva da UFAM.

A ausência da USP como *benchmark* não implica inadequação da instituição. O modelo VRS controla o efeito de escala, e a incompatibilidade entre os vetores de

inputs explica o $\lambda = 0$: a USP opera em um regime de recursos muito distinto. Assim, a DMU Virtual representa um diagnóstico de direção para o presente, não uma prescrição de longo prazo.

Projeção da DMU Virtual e folgas da UFAM

A DMU Virtual permite estimar os *outputs* que a UFAM produziria caso apresentasse a mesma taxa de conversão de recursos observada em seus *benchmarks*. O Quadro 5 apresenta a comparação entre os valores observados e os valores projetados.

Quadro 5 - UFAM: valores observados vs. DMU Virtual

Indicador	UFAM (observado)	DMU Virtual	Diferença absoluta	Diferença relativa
Artigos (Scopus + Pubmed)	33	121	+88	+267%
Patentes (biotecnologia)	14	51	+37	+267%

Fonte: Elaborado pelo autor (2026).

A projeção radial do modelo explica o aumento proporcional:

$$\text{Fator} = \frac{1}{0,2727} = 3,667$$

$$\text{Aumento relativo} = (3,667 - 1) \times 100\% = 266,7\%$$

Além da expansão proporcional, o modelo gera folgas residuais. A UFAM apresenta uma folga de 3,109 artigos PUBMED, indicando necessidade de ajuste adicional mesmo após a projeção radial.

Do lado dos insumos, surgem folgas de 46,186 graduados e 442,400 projetos CAPES, revelando excesso relativo de recursos quando comparada aos *benchmarks*. Esses resultados confirmam que a ineficiência da UFAM é predominantemente de *output*, e não de *input* (Quadro 6).

Quadro 6 - Slacks UFAM

Slacks Input			
DMU	Conc_Grad_Biotec	Bolsas_CAPES	Projetos_CAPES
UFAM	46,186		442,400
Slacks Output			
DMU	SCOPUS	PUBMED	Patentes
UFAM		3,109	

Fonte: Elaborado pelo autor (2026).

Consistência da projeção e interpretação dos resultados

A plausibilidade dos valores projetados pode ser verificada comparando-os com os *outputs* reais das instituições que compõem a DMU Virtual. O Quadro 7 apresenta esses valores.

Quadro 7 - *Outputs* reais dos *benchmarks* da UFAM

IES	Artigos (Scopus + Pubmed)	Patentes
UFV	141	24
UFPB	49	70
UNICAMP	294	149

Fonte: Elaborado pelo autor (2026).

Aplicando os pesos λ :

$$\text{Artigos}_{\text{virtual}} \approx 0,567 \cdot 141 + 0,339 \cdot 49 + 0,094 \cdot 294 \approx 124$$

$$\text{Patentes}_{\text{virtual}} \approx 0,567 \cdot 24 + 0,339 \cdot 70 + 0,094 \cdot 149 \approx 51$$

Os valores obtidos (124 artigos e 51 patentes) são praticamente idênticos aos produzidos pela projeção radial (121 e 51), confirmando a consistência interna do modelo.

Os resultados indicam uma assimetria significativa na taxa de conversão de recursos em *outputs* de inovação entre a UFAM e seus *benchmarks* factíveis. Essa assimetria é produtiva, não estrutural: a UFAM dispõe de recursos comparáveis, mas converte esses recursos em resultados de modo menos eficiente.

Esses números não devem ser interpretados como metas institucionais, nem como juízo de valor sobre a competência dos pesquisadores. A DMU Virtual é um instrumento de diagnóstico, não de prescrição. Ela revela a direção da ineficiência, mas não suas causas, que podem ser culturais, processuais ou estruturais. Os capítulos seguintes, especialmente o Capítulo 3, mostram que o potencial inventivo da UFAM é real e substancial, com 55 ativos identificados, reforçando que o desafio é de conversão, não de capacidade científica.

Escala de eficiência e representação gráfica

A Figura 2, apresentada na Seção 2.4 do artigo-base, é um gráfico de visualização unidimensional e sua leitura requer atenção a três aspectos técnicos que, embora decorram diretamente do modelo DEA, não são explicitados no artigo publicado.

O primeiro aspecto diz respeito aos eixos. O eixo vertical da Figura 2 representa o escore de eficiência, único atributo quantitativo ali representado. O eixo horizontal, por sua vez, não representa variável alguma: as 29 Instituições de Ensino Superior estão distribuídas ao longo desse eixo apenas para fins de legibilidade, sem ordenação ou espaçamento que carreguem significado analítico. A posição horizontal de cada instituição é, portanto, arbitrária, e não deve ser interpretada como proximidade ou distância entre elas. O que o gráfico oferece é uma leitura vertical: cada IES ocupa uma altura específica no eixo y, e é essa altura, e somente ela, que expressa seu desempenho relativo.

O segundo aspecto concerne à fronteira de eficiência. Diferentemente do que ocorre em representações didáticas da DEA, em que a fronteira aparece como uma linha ou superfície, a Figura 2 não exibe a fronteira. Isso decorre da natureza do modelo adotado: a DEA-VRS opera com múltiplos insumos e múltiplos produtos, o que torna a fronteira uma superfície multidimensional, impossível de ser representada em um plano bidimensional. O que o gráfico mostra é apenas a projeção unidimensional dos escores de eficiência, e não a geometria da fronteira. A leitura da posição relativa das IES, portanto, é feita pelo escore, e não pela distância visual em relação a uma linha de fronteira que, no gráfico, não existe.

O terceiro aspecto refere-se à escala visual. Como o modelo foi executado em duas etapas sequenciais, a DEA-VRS na Etapa 1 e a Super-DEA com RUF-Inovação na Etapa 2, os escores das duas etapas encontram-se em escalas distintas: os da Etapa 1 situam-se no intervalo de 0 a 1, enquanto os da Etapa 2 podem ultrapassar 1. Para viabilizar a representação conjunta das 29 IES no mesmo gráfico, aplicou-se a normalização logarítmica, expressa por:

$$Score_{norm} = \frac{\ln(Score + 1)}{\ln(Max + 1)}$$

Essa transformação é monotônica, ou seja, preserva a ordenação original dos escores. Sua função é exclusivamente visual: compatibilizar as escalas das duas etapas e reduzir o efeito de escala do desempenho da USP, permitindo que as demais IES sejam visualmente distinguíveis. Sobre o resultado da normalização, aplicou-se um corte em cinco faixas de cores, associadas a um gradiente de vermelho a verde. Essas faixas são um recurso de leitura gráfica, e não uma escala classificatória independente: a escala subjacente permanece sendo o escore bruto de eficiência, e a única instituição plenamente eficiente continua sendo aquela que atinge o valor 1,0000.

Em síntese, a Figura 2 deve ser lida como um gráfico de visualização unidimensional da eficiência, e não como uma representação geométrica da fronteira DEA. A posição vertical de cada IES reflete seu escore de eficiência; a posição horizontal é meramente distributiva; e as cores resultam de uma normalização logarítmica aplicada para fins de legibilidade. Essa distinção é retomada na Síntese dos Achados, no Capítulo 5, onde o posicionamento da UFAM é descrito tanto pelo escore bruto quanto pela faixa visual correspondente.

CAPÍTULO 3 - POTENCIAL DE PATENTEAMENTO E MATURIDADE TECNOLÓGICA: Análise Integrada com Inteligência Artificial no PPGBIOTEC/UFAM⁸

O presente capítulo é baseado no artigo intitulado "POTENCIAL DE PATENTEAMENTO E MATURIDADE TECNOLÓGICA: Análise Integrada com Inteligência Artificial no PPGBIOTEC/UFAM", mantendo integralmente os resultados e a estrutura do artigo publicado. Adicionalmente, incorpora uma seção complementar (seção 3.7) que aprofunda aspectos metodológicos não tratados no escopo do artigo original, (cujos fundamentos teóricos foram apresentados na Seção 2.8 do REFERENCIAL TEÓRICO (CATEGORIAS PRINCIPAIS) p. 38): O texto base das seções 3.1 a 3.6 é reproduzido integralmente, com adaptações estruturais para adequação ao formato da tese.

Especificamente, a seção 3.7 detalha o processo de validação do modelo *Random Forest* utilizado para classificação automática dos Níveis de Maturidade Tecnológica, apresentando a evolução da Matriz de Confusão e do Coeficiente Kappa de Cohen ao longo de sucessivos refinamentos na rotulagem manual, na calibração de pesos das features e na distribuição amostral. Este aprofundamento, embora não contemplado no artigo por limitações de escopo e espaço editorial, é fundamental para aferir a robustez do método e oferecer subsídios empíricos para a gestão da inovação no PPGBIOTEC/UFAM.

3.1. INTRODUÇÃO: INTELIGÊNCIA ARTIFICIAL APLICADA À AVALIAÇÃO DE PATENTEABILIDADE E MATURIDADE TECNOLÓGICA

O Programa de Pós-Graduação em Biotecnologia da Universidade Federal do Amazonas (PPGBIOTEC/UFAM) desempenha um importante papel na capacitação de pesquisadores, e nas investigações em diversas áreas da biotecnologia. Parte dessas investigações apresenta potencial para aplicações industriais ou para o

⁸ O presente capítulo é uma versão reformatada do artigo intitulado "POTENCIAL DE PATENTEAMENTO E MATURIDADE TECNOLÓGICA: Análise Integrada com Inteligência Artificial no PPGBIOTEC/UFAM", publicado na Revista OWL (OWL Journal), Vol. 4, N. 5, p.1–24, 2026, DOI: 10.5281/zenodo.20114796, (KUME; CHAVES; OLIVEIRA JÚNIOR, 2026). O texto base das seções 3.1 a 3.6 é reproduzido integralmente, conforme normas de autocitação e do modelo tese-artigo, com adaptações estruturais e editoriais para adequação ao formato da tese.

desenvolvimento de novos processos, embora o percurso entre a produção científica e sua eventual proteção por meio de mecanismos formais de propriedade intelectual nem sempre seja linear. A literatura sugere que determinados resultados de pesquisa podem não avançar imediatamente para instrumentos de proteção, aspecto que será analisado ao longo deste trabalho.

Estudos recentes discutem a interação entre universidades e o setor produtivo no Brasil, especialmente em áreas de fronteira tecnológica, como a biotecnologia (Florêncio *et al.*, 2020; Okimoto, 2023); enquanto fatores estruturais/logísticos, típicos da Amazônia, podem influenciar esse processo e afetar o ritmo de desenvolvimento tecnológico (Bejarano *et al.*, 2023; Liboreiro, 2020).

A avaliação da maturidade tecnológica por meio da escala TRL, amplamente utilizada por agências de fomento, permite situar tecnologias ao longo do percurso que vai da pesquisa básica à aplicação industrial. No ambiente acadêmico, é comum que os estudos se concentrem nos níveis iniciais, enquanto avanços para estágios mais elevados exigem validação em ambientes relevantes e maior aproximação com o setor produtivo. Estudos recentes demonstram que técnicas de Processamento de Linguagem Natural (Britt *et al.*, 2008), e aprendizado de máquina podem apoiar a classificação automática de TRL e ampliar a capacidade de análise de grandes volumes de documentos técnicos (Jain; Kumar, 2025).

A articulação entre maturidade tecnológica e potencial de patenteamento oferece uma perspectiva estratégica para compreender como diferentes linhas de pesquisa podem evoluir em direção à inovação. Com base nessa premissa, este estudo realiza uma avaliação integrada desses dois eixos no portfólio de teses e dissertações do PPGBIOTEC/UFAM, utilizando técnicas de NLP e algoritmos de aprendizado de máquina para identificar termos associados à Classificação Internacional de Patentes (CIP) e classificar automaticamente os níveis TRL.

A hipótese central é que, diferentemente da visão tradicional que associa patenteamento apenas a pesquisas aplicadas, o potencial de proteção intelectual no PPGBIOTEC distribui-se de maneira independente do nível de maturidade tecnológica. Considerando a riqueza biotecnológica da Amazônia, pressupõe-se que ativos patenteáveis possam emergir ainda em estágios iniciais, como a prova de conceito (TRL 3), sem exigir necessariamente níveis avançados de desenvolvimento.

3.2. REFERENCIAL TEÓRICO: NLP, CIP, TRL E TRANSFERÊNCIA DE TECNOLOGIA

Processamento de Linguagem Natural (NLP) e análise de patentes

No âmbito da Gestão da Inovação, a habilidade de converter a abundância de dados não estruturados em inteligência estratégica é, gradativamente, um fator de diferenciação decisivo. Devido ao grande volume e à complexidade dos textos, repositórios acadêmicos, teses, dissertações e artigos, são fontes subaproveitadas para a prospecção tecnológica. Nesse sentido, o NLP é o elo entre a linguística e a informática, permitindo que as máquinas entendam a linguagem humana e extraiam informações que um ser humano não conseguiria processar manualmente (Silge; Robinson, 2017).

O uso de NLP para ler e entender documentos de patentes tem se revelado uma estratégia valiosa para identificar tendências, áreas emergentes e o estado da técnica em vários setores tecnológicos (Yoo; Lim; Kim, 2021b). A análise se torna mais clara e objetiva, permitindo identificar inovações e produtos com potencial comercial, quando se consegue extrair de modo ágil e precisa palavras-chave de patentes (Wang; Lin, 2023).

Para transformar o conteúdo qualitativo dos resumos em vetores numéricos, foi imprescindível o fluxo de processamento de NLP utilizado neste estudo. Esse fluxo compreende, inicialmente, a tokenização, processo de segmentação do texto bruto em unidades discretas (*tokens*), como palavras ou termos isolados, permitindo a análise individualizada de seus componentes. Em seguida, procedeu-se à remoção de *stopwords*, palavras funcionalmente comuns, como artigos e preposições, que carecem de carga semântica específica para a distinção de tópicos, e à vetorização, que converte os dados textuais em representações matemáticas. Graças a essa transformação, algoritmos de aprendizado de máquina puderam classificar automaticamente o portfólio do PPGBIOTEC conforme a CIP e os níveis TRL, o que trouxe mais objetividade e reprodutibilidade para a análise.

Classificação Internacional de Patentes (CIP) e o Modelo Biotec-BR

A CIP é gerida pela WIPO, que organiza os ativos de propriedade industrial em uma classificação universal, utilizando um sistema hierárquico de símbolos alfanuméricos (WIPO, 1971). Uma CIP desempenha uma função estratégica ao simplificar, por meio de códigos alfanuméricos, tecnologias que são complexas, facilitando a busca por anterioridades e a prospecção tecnológica (Morales Sánchez *et al.*, 2019). Portanto, identificar com exatidão esses códigos no ambiente acadêmico é fundamental para que a ciência produzida se converta em bens que possam ser protegidos.

Nesse contexto, destaca-se o modelo Biotec-BR, desenvolvido pelo INPI, que agrupa códigos CIP particulares da biotecnologia. De acordo com estudos comparativos realizados por Weid *et al.* (2018), este modelo se evidencia tecnicamente superior às classificações da Organização para a Cooperação e Desenvolvimento Econômico (OCDE) e da própria WIPO, apresentando taxas de recuperação de documentos significativamente mais altas, especialmente na área farmacêutica.

Essa eficácia diagnóstica reforça a necessidade de modelos adaptados a setores estratégicos e justifica a adoção do Biotec-BR como a referência taxonômica deste estudo para a classificação do potencial de patenteamento do portfólio do PPGBIOTEC.

Níveis de maturidade tecnológica (TRL)

A escala TRL foi desenvolvida originalmente pela NASA na década de 1970 como um sistema métrico para avaliar o estágio de maturidade de tecnologias espaciais. A escala varia do nível 1 (princípios básicos observados e reportados) ao nível 9 (sistema efetivamente testado e operacional em ambiente real), permitindo uma avaliação padronizada do desenvolvimento tecnológico (Britt *et al.*, 2008; Quintella, 2017).

O TRL, além de uma métrica de maturidade tecnológica, é amplamente utilizado como ferramenta de gestão de portfólio, permitindo identificar tecnologias emergentes, priorizar investimentos e orientar estratégias de transferência de tecnologia. Sua aplicação em ambientes acadêmicos possibilita visualizar o

posicionamento das pesquisas no espectro ciência-inovação, revelando lacunas e oportunidades.

A adoção deste sistema tem se expandido significativamente para além do setor aeroespacial, sendo atualmente utilizada por agências de fomento, instituições de pesquisa e programas de inovação na Europa e no Brasil como critério para financiamento e avaliação de projetos (Salvador-Carulla *et al.*, 2024). A Comissão Europeia, por exemplo, exige a classificação TRL em propostas de projetos de pesquisa e inovação no âmbito do programa Horizonte Europa.

No contexto acadêmico, a produção científica concentra-se majoritariamente nos estágios iniciais da escala TRL. Pesquisas de base e estudos de prova de conceito tendem a ocupar os níveis 1 a 4, enquanto o desenvolvimento de protótipos e a validação em ambiente relevante situam-se nos níveis 5 a 7. A transição para os níveis 8 e 9, que envolvem sistemas completos e comercialização, é raramente atingida no âmbito de teses e dissertações, demandando parcerias com o setor produtivo e financiamento adicional (Paternostro; Quintella; Leite, 2020).

Para o presente estudo, adotou-se uma adaptação da escala TRL ao contexto da pesquisa acadêmica em biotecnologia, conforme sistematizado no Quadro 1.

Quadro 1 - Critérios para Classificação TRL de Pesquisas Acadêmicas em Biotecnologia

TRL	Designação	Critérios Observáveis em Resumos Acadêmicos
1	Pesquisa básica	Estudos teóricos, revisões de literatura, propostas conceituais, ausência de experimentação
2	Formulação tecnológica	Identificação de compostos, ensaios preliminares <i>in silico</i> , metodologias propostas, caracterização inicial
3	Prova de conceito	Ensaio <i>in vitro</i> , testes de atividade antimicrobiana/antioxidante, caracterização de compostos, resultados preliminares
4	Validação laboratorial	Testes <i>in vivo</i> , modelos animais, desenvolvimento de formulações, ensaios pré-clínicos, resultados quantitativos robustos
5	Validação em ambiente relevante	Escala piloto, testes de campo, estudos de estabilidade, validação pré-clínica avançada, protótipos funcionais
6	Demonstração	Sistema protótipo demonstrado em ambiente relevante, ensaios clínicos iniciais, transferência para setor produtivo
7-9	Sistema completo	Não aplicável ao contexto de teses/dissertações (comercialização)

Fonte: Adaptado de NASA e Britt *et al.* (2008).

Transferência de Tecnologia e Propriedade Intelectual no Contexto Acadêmico

É necessário um sistema estratégico que garanta à descoberta científica proteção legal e valorização comercial, para que a tecnologia gerada na universidade possa ser inserida no mercado. Nesse sentido, a Propriedade Intelectual (PI) é o alicerce desse relacionamento, é a linguagem comum que permite o diálogo entre academia e mercado. Identificar o potencial de proteção, por classificações como a CIP, é o que converte o resultado de uma investigação científica em um ativo protegido, assegurando que os inventores sejam reconhecidos e estimulando o investimento em Pesquisa e Desenvolvimento (P&D) (Fontes, 2020).

A adequada gestão desse patrimônio imaterial é que possibilita obter financiamentos e formar parcerias estratégicas (Bejarano *et al.*, 2023). Entretanto, para que essa transição ocorra de fato, esbarra na formação acadêmica: a disciplina de PI é importante, mas estudos evidenciam que a capacitação atual é insuficiente e desconectada das demandas do mercado (Marinho; Oliveira; Araújo, 2023). Essa falta de formação é, provavelmente, a principal responsável pela disparidade entre a intensa produção científica e a tímida proteção intelectual nos programas de pós-graduação.

O desenvolvimento da biotecnologia no Brasil ocorre majoritariamente nas universidades, o que, somado à pouca interação com o setor privado, impede que a ciência nacional tenha um forte impacto econômico. Para transpor esse obstáculo, é imprescindível fomentar ambientes de inovação aberta, onde o conhecimento circule livremente entre os setores (Florêncio *et al.*, 2020). Em bioeconomia, é imprescindível identificar trajetórias tecnológicas promissoras em seus estágios iniciais para concentrar os esforços de proteção, visto que o sucesso de uma inovação depende, em igual medida, do rigor técnico e da prontidão tecnológica e aceitação social (Hannouf *et al.*, 2025).

Lócus da Pesquisa

O Programa de Pós-Graduação em Biotecnologia da Universidade Federal do Amazonas (PPGBIOTEC/UFAM), criado em 2001, constitui o lócus empírico deste

estudo. O programa oferta cursos de mestrado e doutorado, estruturados em três áreas de concentração: Biotecnologias para a Área Agroflorestal, Biotecnologias para a Saúde e Gestão da Inovação em Biotecnologia. No período analisado (2015-2024), o programa registrou 259 defesas de teses e dissertações, cujos resumos foram coletados por meio do repositório institucional TEDE/UFAM. Uma caracterização mais abrangente do ecossistema de inovação da UFAM, incluindo o papel da Pró-Reitoria de Inovação Tecnológica (PROTEC).

3.3. PROCEDIMENTOS METODOLÓGICOS: FLUXO DE INTELIGÊNCIA DE DADOS PARA AVALIAÇÃO DE PATENTEABILIDADE E MATURIDADE

Dada a necessidade de alinhar a produção acadêmica à maturidade tecnológica discutida anteriormente, a presente seção detalha o fluxo de inteligência de dados desenvolvido para este diagnóstico.

Delineamento metodológico da pesquisa

O estudo caracteriza-se como uma investigação exploratório-descritiva, de abordagem qualiquantitativa, voltada à análise documental das teses e dissertações defendidas no PPGBIOTEC/UFAM entre 2015 e 2024. Foram empregadas técnicas de NLP e algoritmos de aprendizado de máquina (ML) para identificar o potencial de patenteamento e classificar a maturidade tecnológica (TRL) das pesquisas.

Bases de dados e coleta

Os dados foram obtidos do repositório institucional TEDE/UFAM, por meio do protocolo OAI-PMH, seguindo diretrizes de reprodutibilidade e Ciência Aberta (Mojjada, 2025). Um script em R (R Core Team, 2026) coletou automaticamente os metadados dos trabalhos vinculados ao PPGBIOTEC, incluindo título, autoria, orientação, ano, tipo de documento e resumo. O modelo Biotec-BR, elaborado pelo INPI, foi adotado como referência para a Classificação Internacional de Patentes (CIP).

Pré-processamento dos textos

Os resumos foram submetidos a padronização para minúsculas, remoção de pontuação e *stopwords*, tokenização e formação de bigramas, preservando termos técnicos compostos característicos da biotecnologia. As rotinas foram implementadas com as bibliotecas *tm*, *stringr* e *dplyr*.

Identificação do potencial de patenteamento segundo a CIP

O potencial de patenteamento foi identificado por meio de correspondência textual entre os resumos e um dicionário de termos associados aos códigos CIP do modelo Biotec-BR. A presença de ao menos um termo correlato foi considerado indicativo de potencial de patenteamento, em linha com práticas de prospecção tecnológica.

Classificação da maturidade tecnológica (TRL)

Rotulagem manual e construção de *corpus* de treinamento

Foi construído um *corpus* de 80 resumos rotulados manualmente por especialistas, com base em critérios ontológicos padronizados (Salvador-Carulla *et al.*, 2024). A literatura aponta que, em cenários de *small datasets*, a qualidade da amostragem e da rotulagem é mais relevante que o volume absoluto (Han; Williamson; Fong, 2021; Lee *et al.*, 2025). O uso de NLP para extração de inteligência tecnológica em documentos técnicos é igualmente validado por (Wang; Lin, 2023).

Vetorização

Após o pré-processamento, os textos foram vetorizados pelo método *Bag-of-Words* utilizando *text2vec*, com criação do vocabulário, remoção de termos raros e geração da matriz termo-documento.

Treinamento e validação

O modelo *Random Forest* foi treinado com as bibliotecas *parsnip* e *ranger*, utilizando particionamento estratificado (80% treino; 20% validação) e 100 árvores. A escolha do algoritmo se justifica por sua robustez em alta dimensionalidade e bom desempenho em classificação de textos (Breiman, 2001). A acurácia foi utilizada para medir a proporção de classificações corretas, enquanto a matriz de confusão permitiu identificar, detalhadamente, quais classes o modelo acertou e onde ocorreram erros.

Classificação da base completa

O modelo treinado foi aplicado ao restante da base, gerando previsões de TRL e escores de confiança. Para fins analíticos, os níveis foram agrupados em duas faixas, conforme Salvador-Carulla *et al.* (2024): TRL 1-3 (pesquisa básica) e TRL 4-6 (desenvolvimento).

Integração dos resultados e análises estatísticas

Os resultados de TRL e CIP foram integrados em uma única base, permitindo: (i) analisar a distribuição do potencial de patenteamento por faixa TRL; (ii) identificar pesquisas com maior probabilidade de conversão tecnológica (TRL ≥ 3 e confiança $> 0,6$); e (iii) calcular os coeficientes de correlação de Pearson e Spearman entre maturidade tecnológica e potencial de patenteamento. As análises foram conduzidas em R com *dplyr*, *tidyr*, *corrplot* e *ggplot2*.

Crítérios de inclusão e exclusão

Foram incluídas todas as teses e dissertações do PPGBIOTEC/UFAM com resumos disponíveis em português no período de 2015 a 2024. Registros duplicados ou sem resumo foram excluídos. Para a análise integrada, todos os documentos com classificação TRL foram considerados, independentemente da presença de potencial de patenteamento.

Limitações metodológicas do estudo

A análise baseia-se em resumos, o que pode omitir detalhes experimentais relevantes. A classificação TRL depende da qualidade da rotulagem manual e da natureza subjetiva da linguagem acadêmica. Ainda assim, fluxos rigorosos de pré-processamento e modelos supervisionados têm se mostrado adequados para análises exploratórias em biotecnologia (Silge; Robinson, 2017; Wang; Lin, 2023).

Bibliotecas R utilizadas na implementação

A implementação foi realizada em RStudio (Posit), ambiente alinhado aos princípios da Ciência Aberta. As bibliotecas empregadas contemplam desde pré-processamento textual (*tm*, *stringr*), vetorização (*text2vec*), modelagem (*parsnip*, *ranger*, *rsample*) e visualização (*ggplot2*, *plotly*), até a construção da interface interativa (*shiny*, *shinydashboard*). A adoção de ferramentas livres e de código aberto reforça o compromisso com transparência e reprodutibilidade.

Quadro 2 - Apresenta as principais bibliotecas utilizadas, suas respectivas funções e justificativas para aplicação no contexto deste estudo.

Biblioteca	Função Principal	Justificativa / Aplicação no Estudo
<i>shiny</i>	Desenvolvimento de aplicativos web interativos	Interface interativa para explorar dados, aplicar filtros e visualizar resultados sem programação.
<i>shinythemes</i>	Temas prontos para <i>Shiny</i>	Melhora visual do app, tornando-o mais profissional e intuitivo.
<i>shinydashboard</i>	<i>Layouts dashboard</i> estilo	Criação de painéis estatísticos e <i>cards</i> de indicadores-chave.
<i>readxl</i>	Leitura de arquivos Excel	Importa bases de dados em .xlsx (teses, dissertações, classificação CIP).
<i>openxlsx</i>	Escrita de arquivos Excel	Gera relatórios consolidados em Excel com múltiplas abas.
<i>janitor</i>	Limpeza e padronização de nomes de colunas	Padroniza nomes de colunas para consistência e evitar erros.
<i>dplyr</i>	Manipulação e transformação de dados	Filtra, agrupa, sumariza e junta bases ao longo da análise.
<i>tidyr</i>	Organização de dados (formato longo/wide)	Prepara dados para visualizações, como matriz TRL x Patente.
<i>stringr</i>	Manipulação de <i>strings</i> (texto)	Limpa resumos e busca correspondência de termos CIP.

Biblioteca	Função Principal	Justificativa / Aplicação no Estudo
<i>tm</i>	<i>Text Mining</i> (mineração de texto)	Fornece <i>stopwords</i> em português e auxilia na limpeza textual.
<i>tidytext</i>	Análise textual	Base conceitual para estruturar textos.
<i>text2vec</i>	Vetorização de texto e criação de matrizes	Converte resumos em representações numéricas (bag-of-words) para classificação TRL.
<i>parsnip</i>	Interface unificada para modelos de <i>machine learning</i>	Abstração para treinar <i>Random Forest</i> com sintaxe padronizada.
<i>ranger</i>	Implementação rápida de <i>Random Forest</i> em R	Classifica níveis TRL a partir de dados vetorizados (rápido e robusto).
<i>rsample</i>	Amostragem e particionamento de dados	Divide amostra rotulada em treino (80%) e validação (20%), estratificada por TRL.
<i>ggplot2</i>	Criação de gráficos estáticos	Gera gráficos exploratórios e figuras para o artigo (distribuições, correlações).
<i>plotly</i>	Gráficos interativos	Cria gráficos dinâmicos no <i>Shiny</i> , como matriz TRL x Patente.
<i>DT</i>	Tabelas interativas	Exibe tabelas no <i>Shiny</i> com ordenação, busca e paginação.
<i>corrplot</i>	Visualização de matrizes de correlação	Visualiza correlações entre variáveis na análise exploratória.

Fonte: Elaborado pelo autor (2026).

Considerações éticas e de escopo

Por tratar-se de análise documental de dados públicos, não houve necessidade de submissão a Comitê de Ética. Os resultados são apresentados agregadamente, sem identificação individual, resguardando potenciais estratégias futuras de proteção intelectual.

A implementação foi realizada em RStudio, atualmente Posit (POSIT, 2022), plataforma desenvolvida por uma *Public Benefit Corporation* (PBC) (LAW INSIDER, 2025), alinhada a princípios de transparência e responsabilidade social. Esse enquadramento institucional converge com os fundamentos da Ciência Aberta, uma vez que todas as ferramentas utilizadas são livres, auditáveis e amplamente validadas pela comunidade científica. A aplicação construída em Shiny não se configura como produto de *software* comercial, mas como instrumento científico desenvolvido especificamente para este estudo, desempenhando papel análogo ao de equipamentos laboratoriais: uma ferramenta para explorar padrões, revelar potenciais ocultos e subsidiar decisões estratégicas no âmbito acadêmico.

3.4. RESULTADOS: DISTRIBUIÇÃO TRL, POTENCIAL DE PATENTEAMENTO E ATIVOS ESTRATÉGICOS

Caracterização da base de dados

O processamento dos 259 registros revelou um ecossistema de pesquisa fortemente concentrado nas etapas iniciais de desenvolvimento científico. A aplicação do modelo *Random Forest* (treinado com 80 amostras) permitiu classificar a totalidade do *corpus* com uma confiança média de 0,89 para o grupo de pesquisa básica.

Quadro 3 - Distribuição de Frequência por Faixa de Maturidade Tecnológica (TRL)

TRL	Classificação	n	% total	Confiança Média
2	Conceito Tecnológico	7	2,7%	1,00 ⁹
3	Prova de Conceito	238	91,9%	0,89
4	Validação em Laboratório	12	4,6%	1,00
5	Validação em Ambiente Relevante	1	0,4%	1,00
6	Piloto em Campo	1	0,4%	1,00
Total		259	100%	0,98

Fonte: Elaborado pelo autor (2026).

A predominância de TRL 2-3 (94,6%) é característica de programas de pós-graduação *stricto sensu*, cujo foco reside na descoberta de novos mecanismos, no isolamento de moléculas e em provas de conceito laboratoriais. No entanto, a identificação de 14 ativos (5,4%) em níveis de desenvolvimento (TRL 4-6) sinaliza a existência de tecnologias que já superaram a fase de bancada e possuem protocolos de validação mais robustos.

Potencial de Patenteamento

A análise de aderência ao modelo Biotec-BR identificou que o potencial de patenteamento está presente em uma parcela significativa do programa, superando a percepção inicial de baixa aplicabilidade.

- Total de pesquisas com potencial de patente: 55 registros;
- Taxa de conversão potencial: 21,2% do portfólio.

⁹ Registra-se uma cautela metodológica: a confiança máxima (1,00) observada nas faixas TRL 4-6 decorre da predominância, nesses níveis, de registros rotulados manualmente que compuseram o *corpus* de treinamento, o que pode superestimar a acurácia do modelo para essas classes.

O Quadro 4 cruza a maturidade tecnológica com a presença de códigos CIP, revelando a distribuição do potencial de inovação:

Quadro 4 - Análise Integrada: Maturidade (TRL) vs. Potencial de Patenteamento

Faixa TRL	Total de Pesquisas (n)	Pesquisas com Potencial de Patente	Taxa de Patenteabilidade Interna
1-3 (Básica)	245	49	20,00%
4-6 (Desenvolvimento)	14	6	42,90%
Total	259	55	21,20%

Fonte: Elaborado pelo autor (2026).

Observa-se que, embora a maior quantidade absoluta de patentes potenciais conste na pesquisa básica (n=49), a probabilidade de uma pesquisa ter potencial de patente é mais que o dobro nas pesquisas de nível TRL 4-6 (42,9% contra 20,0%).

Análise de Correlação e Teste de Hipótese

Contrariando a expectativa inicial de uma relação linear forte, os testes estatísticos revelaram que a maturidade tecnológica e o potencial de patenteamento são fenômenos que caminham independentemente no PPGBIOTEC.

- Coeficiente de Pearson: 0,1318 (não há evidência de correlação linear);
- Coeficiente de Spearman: 0,1505.

Com isso, sugere-se que, no PPGBIOTEC, a TRL isoladamente não é um indicador absoluto do potencial de patenteabilidade. Uma pesquisa de TRL 2 (formulação tecnológica) pode apresentar um potencial de proteção intelectual tão nítido quanto um protótipo TRL 5, especialmente no que tange ao isolamento de novos compostos e sequenciamentos genéticos originais.

Identificação de ativos potenciais e recomendações estratégicas

A etapa final da análise integrada consistiu na filtragem dos ativos com maior probabilidade de sucesso em uma eventual jornada de transferência de tecnologia. Foram selecionadas 55 pesquisas que atenderam simultaneamente a três critérios de rigor: (i) correspondência positiva com códigos CIP; (ii) maturidade tecnológica TRL $\geq$ 3; e (iii) confiança de predição do modelo superior a 0,6.

A listagem desses ativos, com a identificação parcialmente omitida para preservar a integridade da estratégia de proteção intelectual da UFAM, é apresentada no Quadro 5.

Quadro 5 - Portfólio de Ativos com Potencial de Inovação Identificados (N=55)

ID	Ano	CIP	Relação com CIP
43XX	2014	B09C1/10	Remediação solo
61XX	2015	C12P*	Bioprodução
70XX	2018	C12P*	Bioprodução
73XX	2019	B09C1/10	Remediação solo
73XX	2019	C12P*	Bioprodução
83XX	2021	C12P*	Bioprodução
42XX	2014	B09C1/10	Remediação solo
50XX	2015	C12P*; B82Y5/00	Nanotecnologia
50XX	2015	B09C1/10	Remediação solo
53XX	2014	C12P*	Bioprodução
53XX	2014	G01N33/53*; G01N33/54*	Imunoensaios
56XX	2015	C12P*	Bioprodução
57XX	2017	A61K48*; B09C1/10	Terapia gênica
57XX	2016	C12P*	Bioprodução
57XX	2016	C12P*	Bioprodução
59XX	2017	A61K8/99	Cosmética
59XX	2017	A61K8/99	Cosmética
62XX	2017	G01N33/53*; G01N33/54*	Imunoensaios
62XX	2017	C12P*	Bioprodução
62XX	2015	B09C1/10	Remediação solo
66XX	2014	C12P*	Bioprodução
66XX	2015	A61K8/99	Cosmética
66XX	2017	B09C1/10	Remediação solo
67XX	2018	C12P*	Bioprodução
67XX	2017	B09C1/10	Remediação solo
69XX	2015	C12P*	Bioprodução
69XX	2017	C12P*	Bioprodução
70XX	2019	B09C1/10	Remediação solo
70XX	2019	A61K8/99	Cosmética
73XX	2019	C12P*	Bioprodução
72XX	2019	C12P*	Bioprodução
76XX	2018	B09C1/10	Remediação solo
79XX	2020	B09C1/10	Remediação solo
84XX	2018	A61K49/14; A61K49/16	Bioimagem
84XX	2021	C12P*	Bioprodução
84XX	2021	C12P*	Bioprodução
85XX	2021	C12P*; A23C9/12*	Laticínios
89XX	2018	C12P*	Bioprodução
85XX	2021	B09C1/10	Remediação solo
87XX	2021	C12P*; A23L2/84	Bebidas
88XX	2021	A23L2/84	Bebidas
89XX	2022	C12P*	Bioprodução
90XX	2022	C12P*	Bioprodução

ID	Ano	CIP	Relação com CIP
91XX	2022	C12P*	Bioprodução
91XX	2022	A61K8/99	Cosmética
92XX	2022	A61K49/14; A61K49/16	Multipropósito
92XX	2022	C12P*	Bioprodução
94XX	2023	A61K8/99	Cosmética
96XX	2023	A61K8/99	Cosmética
97XX	2023	A23L2/84	Bebidas
97XX	2023	A61K8/99; A61K49/14	Cosmética
98XX	2023	C12P*	Bioprodução
99XX	2023	B09C1/10	Remediação solo
10XXX	2024	A61K8/99	Cosmética
10XXX	2024	A23L2/84	Bebidas

Fonte: Elaborado pelo autor (2026).

Dentro deste grupo, destacam-se registros que alcançaram confiança máxima (1,0), indicando uma assinatura terminológica clara tanto de potencial inventivo quanto de prontidão tecnológica. Exemplos notáveis incluem pesquisas voltadas para o desenvolvimento de bioprocessos com fungos amazônicos (C12P) e formulações de nanopartículas de prata com efeito medicinal (B82Y), áreas onde o PPGBIOTEC demonstra consolidada competência técnica.

A baixa correlação observada entre TRL e Patente (0,1318) reforça a necessidade de uma vigilância tecnológica constante. No cenário da biotecnologia amazônica, a "janela de patenteabilidade" abre-se muito cedo. Esperar que uma pesquisa atinja níveis elevados de TRL (7-9) para iniciar o processo de proteção intelectual pode resultar na perda de prioridade frente a competidores internacionais que monitoram a biodiversidade da região.

Em suma, o uso de NLP e ML demonstrou ser uma ferramenta eficaz para dar visibilidade a ativos que, de outra forma, permaneceriam "ocultos" no volume documental do repositório institucional. Com isso, o PPGBIOTEC/UFAM deixa de ser um gestor reativo e atua na prospecção tecnológica proativa, significando melhor alocação de recursos e maior velocidade no impacto socioeconômico da ciência gerada na Amazônia.

3.5. DISCUSSÃO: IMPLICAÇÕES PARA A INOVAÇÃO NA AMAZÔNIA E PARA AS DINÂMICAS INSTITUCIONAIS DA UFAM RELACIONADAS À INOVAÇÃO

A identificação de 55 pesquisas com potencial de patente, correspondendo a 21,2% do portfólio analisado, revela um cenário de inovação muito mais promissor do que o sugerido pela ausência de depósitos na base internacional PATENTSCOPE. Esse descompasso evidencia a existência de um “vale da morte” institucional (Salvador-Carulla *et al.*, 2024), no qual descobertas com valor inventivo não avançam para mecanismos formais de proteção. A taxa encontrada é expressiva para um ambiente acadêmico e indica que o programa produz, consistentemente, conhecimento com aderência tecnológica ao modelo Biotec-BR.

A literatura sobre transferência de tecnologia em ICTs públicas (De Menezes *et al.*, 2025; Santos *et al.*, 2024), aponta que a superação do chamado “vale da morte” institucional, isto é, a dificuldade de converter conhecimento acadêmico em inovação protegida e transferida, costuma envolver três funções analíticas inter-relacionadas: (i) *scouting*, ou a identificação sistemática de ativos tecnológicos latentes; (ii) proteção estratégica, compreendendo mecanismos de blindagem precoce da invenção, como depósitos provisórios ou patentes de processo; e (iii) validação e *matchmaking*, que abrange a elevação do nível de maturidade tecnológica (TRL) por meio de editais de escalonamento e a conexão com potenciais parceiros industriais. Os 55 ativos identificados neste estudo (Quadro 5) constituem um conjunto sobre o qual tais funções poderiam ser exercidas, em linha com o que a literatura recomenda para contextos semelhantes. A distribuição desses ativos entre os perfis de menor (TRL 3, n=49) e maior maturidade (TRL 4-6, n=6) sugere que diferentes ênfases seriam adequadas, embora a decisão sobre sua implementação pertença à esfera institucional, e não a este estudo.

A análise da maturidade tecnológica mostra que 94,6% das pesquisas situam-se nos níveis TRL 1-3, refletindo a vocação do PPGBIOTEC para a pesquisa de fronteira. Esse perfil não representa fragilidade, mas sim a natureza da investigação biotecnológica, que se concentra na descoberta de novos compostos, mecanismos e processos. Ainda que minoritária, a presença de 5,4% das pesquisas nos níveis TRL 4-6 demonstra que parte da produção já avança para etapas de validação laboratorial

e testes em ambiente relevante, especialmente em temas associados à biodiversidade amazônica.

Um dos achados mais significativos é a baixa correlação entre maturidade tecnológica e potencial de patenteamento (Pearson = 0,1318). Esse resultado refuta a ideia de que a proteção intelectual depende necessariamente de níveis avançados de desenvolvimento. No contexto da biotecnologia amazônica, o valor inventivo pode emergir ainda em fases iniciais, como na identificação de moléculas bioativas ou no desenvolvimento de processos fermentativos, confirmando a existência de uma janela de oportunidade precoce para proteção intelectual. Essa dinâmica está alinhada às observações de Wang; Lin (2023), que defendem a necessidade de *roadmapping* tecnológico capaz de identificar ativos protegíveis ainda na fase de prova de conceito.

A consolidação de 55 ativos de alta fidelidade, resultantes da combinação entre $TRL \geq 3$ e confiança de predição superior a 0,6, demonstra que o programa possui massa crítica suficiente para estruturar um portfólio estratégico de inovação. Além do volume, a diversidade temática distribuída ao longo da última década indica maturidade institucional e continuidade na produção de conhecimento aplicável, especialmente em áreas como fermentações, bioprospecção e farmacogenômica.

Esses resultados têm implicações diretas para a política institucional de inovação. A baixa correlação entre maturidade e patenteamento sugere que resultados passíveis de proteção podem ocorrer bem antes de uma tecnologia estar plenamente desenvolvida. Esse padrão indica que as estratégias institucionais de prospecção tecnológica podem se beneficiar de métodos que não se restrinjam apenas a pesquisas que estão próximas da aplicação industrial, mas que também incluam investigações em estágios mais iniciais.

As limitações do estudo devem ser consideradas na interpretação dos resultados. A análise concentrou-se em um único programa de pós-graduação, restringindo a generalização para outros contextos. A identificação do potencial de patenteamento depende da presença explícita de termos nos resumos, podendo subestimar pesquisas descritas de maneira mais genérica. Além disso, embora a ampliação da amostra rotulada tenha fortalecido o modelo de classificação TRL, a linguagem acadêmica continua impondo desafios à predição automática. Para estudos futuros, recomenda-se ampliar o *corpus* para incluir textos completos, acompanhar longitudinalmente os ativos identificados, aplicar a metodologia a outros

programas de biotecnologia e testar modelos semânticos avançados, como BERTimbau e BioBERT, que podem oferecer ganhos de precisão.

3.6. CONSIDERAÇÕES FINAIS: SÍNTESE DO POTENCIAL DE INOVAÇÃO DO PPGBIOTEC

Este estudo apresentou uma análise integrada do potencial de patenteamento e da maturidade tecnológica das pesquisas desenvolvidas no PPGBIOTEC/UFAM entre 2015 e 2024, utilizando técnicas de NLP e aprendizado de máquina aplicadas a 259 teses e dissertações. Os resultados revelam que 21,2% das pesquisas apresentam potencial de patenteamento, um volume expressivo que contrasta com a ausência de depósitos na base PATENTSCOPE e evidência um hiato crítico entre a geração de conhecimento inventivo e sua proteção formal.

A distribuição das pesquisas nos níveis TRL 1-3 (94,6%) confirma a vocação do programa para a pesquisa de fronteira, enquanto os 5,4% que alcançam TRL 4-6 demonstram esforços de validação tecnológica. A baixa correlação entre maturidade e patenteamento (Pearson = 0,1318) indica que a inovação protegível emerge tanto da pesquisa básica quanto da aplicada, abrindo uma janela de oportunidade precoce para proteção intelectual. A consolidação de 55 ativos de alta fidelidade reforça a existência de um núcleo estratégico de tecnologias com potencial imediato para ações de redação de patentes e prospecção de parcerias industriais.

A metodologia baseada em NLP mostrou-se eficaz para transformar um repositório documental extenso em um mapa de inteligência tecnológica, reduzindo assimetrias de informação e apoiando a gestão da inovação. Os achados desafiam a noção de que a proteção intelectual deve ocorrer apenas em estágios avançados de desenvolvimento e sugerem que políticas institucionais de fomento à inovação devem monitorar ativamente a pesquisa básica, onde residem ativos biotecnológicos de alto valor.

Por fim, o estudo reforça a urgência de estratégias que superem o “vale da morte” identificado. A conversão do talento científico do PPGBIOTEC em impacto socioeconômico depende de suporte especializado à proteção intelectual e do fortalecimento das interações universidade-empresa. O panorama apresentado oferece subsídios concretos para que a biodiversidade amazônica seja protegida e

convertida em inovações capazes de impulsionar o desenvolvimento regional sustentável.

3.7. ANÁLISE COMPLEMENTAR: VALIDAÇÃO DO MODELO TRL E EVOLUÇÃO DAS MÉTRICAS DE DESEMPENHO

Esta seção detalha a validação do modelo *Random Forest* para a classificação automática dos TRL, abordando aspectos metodológicos não incluídos no artigo. São descritos o processo de rotulagem manual aprimorado, a evolução da matriz de confusão e do coeficiente Kappa de Cohen ao longo de sucessivos refinamentos, bem como as decisões metodológicas adotadas para otimizar o desempenho do classificador.

Construção do Corpus, Estratégia de Rotulagem e Arquitetura do Modelo

O corpus de treinamento foi composto por 65 resumos de teses e dissertações do PPGBIOTEC/UFAM. Após análise preliminar dos resultados e identificação de confusão sistemática entre classes adjacentes, especialmente TRL 2 ↔ TRL 3 e TRL 4 ↔ TRL 5, optou-se por uma estratégia de agrupamento em três macro-níveis, preservando a semântica da escala original, conforme apresentado no Quadro 6.

Quadro 6 – Agrupamento em Níveis TRL

Grupo	TRLs Originais	TRL Final	Descrição
Baixo	1, 2	1	Pesquisa básica e caracterização inicial
Médio	3, 4	3	Validação experimental e laboratorial
Alto	5	5	Validação em escala relevante

Fonte: Elaborado pelo autor (2026).

A rotulagem manual foi realizada pelo autor, seguindo critérios baseados na literatura sobre adaptação da escala TRL para contextos acadêmicos (Falvo; Weidener; Karlsson, 2025; Paternostro; Quintella; Leite, 2020; Salvador-Carulla *et al.*, 2024), bem como em diretrizes para classificação de estágios de maturidade em projetos de pesquisa (Falvo; Weidener; Karlsson, 2025). A distribuição final do corpus encontra-se no Quadro 7.

Quadro 7 - Distribuição Final do Corpus de Validação (n=65)

TRL Final	Quantidade	Percentual
1 (Baixo)	26	40%
3 (Médio)	30	46%
5 (Alto)	9	14%
Total	65	100%

Fonte: Elaborado pelo autor (2026).

As 11 features foram definidas com base em critérios observáveis diretamente nos resumos acadêmicos, conforme literatura especializada na adaptação da escala TRL para contextos de pesquisa universitária (Falvo; Weidener; Karlsson, 2025; Salvador-Carulla *et al.*, 2024). Cada *feature* corresponde a um indicador linguístico de determinado estágio de maturidade tecnológica. Os pesos foram calibrados por meio de otimização automática (*grid search*), técnica que testa sistematicamente combinações de parâmetros para identificar a configuração que maximiza o desempenho do modelo, conforme registrado no Quadro 8.

Quadro 8 - Features e Pesos Finais após Otimização

feature_name	trl_1	trl_3	trl_5
tem_pesquisa_basica	0,515	0,712	0,562
tem_caracterizacao	0,427	0,794	0,818
tem_expressao_proteica	0,463	0,768	0,175
tem_otimizacao	0,448	0,394	0,446
tem_escala_piloto	0,516	0,511	0,455
tem_validacao_in_vivo	0,806	0,713	0,678
tem_prototipo	0,688	0,693	0,512
tem_aplicacao_industrial	0,748	0,538	0,614
tem_analise_viabilidade	0,607	0,417	0,335
eh_doutorado	0,607	0,600	0,640
palavras_no_resumo	0,933	0,293	0,264

Fonte: Elaborado pelo autor (2026).

A atribuição de pesos diferenciados segue a recomendação de Falvo, Weidener e Karlsson (2025), segundo os quais cada critério deve refletir sua importância relativa em cada nível TRL. Para operacionalizar essa recomendação, foi construído um dicionário de keywords que mapeia termos técnicos e científicos a cada uma das 11 features. A cada keyword associou-se um peso (variando de 0 a 1), que indica a força do indicador linguístico para a ativação da respectiva feature no resumo. Esse dicionário, apresentado no Quadro 9, permite que o modelo converta a linguagem natural dos resumos em vetores estruturados de features ponderadas.

Quadro 9 - Dicionário de Keywords e Pesos por Feature

feature_name	keyword	peso
tem_pesquisa_basica	prospecção	1.0
tem_pesquisa_basica	screening	1.0
tem_pesquisa_basica	isolamento	1.0
tem_pesquisa_basica	identificação	1.0
tem_pesquisa_basica	screening	1.0
tem_pesquisa_basica	bioprospecção	1.0
tem_pesquisa_basica	rna-seq	1.0
tem_pesquisa_basica	transcriptoma	1.0
tem_pesquisa_basica	genoma	1.0
tem_caracterizacao	microscopia	0.8
tem_caracterizacao	western blot	0.8
tem_caracterizacao	pcr	0.6
tem_caracterizacao	caracterização	1.0
tem_caracterizacao	análise	1.0
tem_caracterizacao	perfil químico	1.0
tem_caracterizacao	espectrometria	0.8
tem_caracterizacao	cromatografia	0.8
tem_expressao_proteica	clonagem	1.0
tem_expressao_proteica	expressão	1.0
tem_expressao_proteica	expressão heteróloga	1.0
tem_expressao_proteica	proteína recombinante	1.0
tem_expressao_proteica	pichia pastoris	1.0
tem_expressao_proteica	escherichia coli	0.8
tem_otimizacao	otimização	1.0
tem_otimizacao	delineamento experimental	1.0
tem_otimizacao	superfície de resposta	1.0
tem_otimizacao	planejamento fatorial	1.0
tem_escala_piloto	biorreator	1.0
tem_escala_piloto	escala piloto	1.0
tem_escala_piloto	fermentação submersa	0.8
tem_escala_piloto	escalonamento	1.0
tem_validacao_in_vivo	camundongos	1.0
tem_validacao_in_vivo	modelo animal	1.0
tem_validacao_in_vivo	ensaio <i>in vivo</i>	1.0
tem_prototipo	protótipo	1.0
tem_prototipo	sistema funcional	1.0
tem_prototipo	plataforma	0.7
tem_aplicacao_industrial	aplicação industrial	1.0
tem_aplicacao_industrial	patente	1.0
tem_aplicacao_industrial	mercado	1.0
tem_aplicacao_industrial	produto comercial	1.0
tem_analise_viabilidade	viabilidade econômica	1.0
tem_analise_viabilidade	análise de custos	1.0
tem_analise_viabilidade	retorno de investimento	1.0

Fonte: Elaborado pelo autor (2026).

Evolução das Métricas, Matriz de Confusão e Interpretação Estatística

Ao longo do processo de refinamento, foram testadas diferentes configurações de pesos, distribuições amostrais e estratégias de agrupamento. Cada configuração foi avaliada segundo os critérios de acurácia e coeficiente Kappa, permitindo comparar o desempenho relativo entre iterações. A evolução das principais métricas encontra-se no Quadro 10, que evidencia melhora progressiva do modelo, culminando na configuração final com três classes (1, 3 e 5) e pesos otimizados.

Quadro 10 - Evolução da Acurácia e do Kappa

Iteração	Configuração	Amostras	Acurácia	Kappa
1	5 classes, pesos iniciais	39	33,33%	0,129
2	5 classes, distribuição natural	59	18,64%	-0,059
3	3 classes (1,2,3), pesos ajustados	60	56,67%	0,258
4	3 classes (1,3,5), pesos otimizados	65	67,69%	0,421

Fonte: Elaborado pelo autor (2026).

A matriz de confusão final, apresentada no Quadro 11, mostra que TRL 3 (Médio) apresentou desempenho perfeito (100% de acerto), enquanto TRL 1 (Baixo) obteve 50% de acerto e TRL 5 (Alto) apresentou desempenho limitado (11%), reflexo do reduzido número de amostras ($n=9$) e da similaridade lexical com descrições de validação laboratorial.

Quadro 11 - Matriz de Confusão Final ($n=65$)

Real \ Predito	1	3	5	Total	Acurácia
1	13	13	0	26	50,00%
3	0	30	0	30	100,00%
5	2	6	1	9	11,10%
Total	15	49	1	65	67,69%

Fonte: Elaborado pelo autor (2026).

O coeficiente Kappa final foi de 0,4206. Para avaliar a precisão dessa estimativa, utilizou-se o método estatístico de *bootstrap* (Markus; Groenen, 1998), uma técnica de reamostragem com 500 replicações a partir das 65 amostras rotuladas, permitindo estimar o intervalo de confiança do Kappa sem assumir distribuição normal dos dados. O intervalo de confiança de 95% obtido foi [0,2687;

0,5949], estatisticamente significativo por não conter o valor zero. Segundo a escala de Landis e Koch (1977), o classificador situa-se entre “concordância razoável” e “concordância moderada”, aproximando-se da faixa “substancial” no limite superior. Considerando as peculiaridades do corpus — textos acadêmicos de biotecnologia com alta similaridade linguística — o resultado é robusto e validado pela literatura especializada (Congalton, 1991; Foody, 2002).

A análise dos erros revelou três padrões principais, apresentados no Quadro 12, indicando que a distinção linguística entre pesquisa básica (TRL 1–2) e validação laboratorial (TRL 3–4) é tênue, assim como a diferença entre validação laboratorial e validação em escala (TRL 5). A confusão entre TRL 1 e TRL 5, embora rara, é metodologicamente relevante, pois resumos de polos opostos podem compartilhar vocabulário quando descrevem “potencial aplicativo”.

Quadro 12 – Padrões de Erro

Padrão de Erro	Ocorrências	Interpretação
TRL 1 → TRL 3	13	Pesquisa básica classificada como validação laboratorial (menção a "aplicação" ou "validação" no texto)
TRL 5 → TRL 3	6	Validação em escala classificada como laboratorial (descrição em termos de bancada)
TRL 5 → TRL 1	2	Validação em escala classificada como pesquisa básica (vocabulário de descoberta inicial)

Fonte: Elaborado pelo autor (2026).

Complementarmente à validação do classificador TRL, e em resposta à necessidade de sustentar estatisticamente a análise de correlação apresentada na Seção 3.4.3 do artigo-base, aplicou-se o teste t de *Student* (gl = 257) aos coeficientes de Pearson e Spearman. O coeficiente de Pearson ($r = 0,1318$) apresentou p-valor = 0,0340, com intervalo de confiança de 95% [0,0100; 0,2496], sendo estatisticamente significativo ($p < 0,05$). O coeficiente de Spearman ($\rho = 0,1505$) apresentou p-valor = 0,0154, também estatisticamente significativo ($p < 0,05$).¹⁰ Embora ambos sejam significativos, a magnitude é muito fraca ($r^2 = 0,0174$; menos de 2% de variância compartilhada), o que sustenta com rigor estatístico a conclusão de que TRL e potencial de patenteamento são fenômenos praticamente independentes no PPGBIOTEC.

¹⁰ Como a variável potencial_patente é binária, há empates nos *ranks*, e o p-valor de Spearman foi calculado por aproximação assintótica (método implementado no R via *cor.test*).

Quadro 4 - Coeficientes de correlação e testes de significância (N = 259)

Método	Coeficiente	p-valor	IC 95%	r ²	gl	Significância
Pearson	r = 0,1318	0,0340	[0,0100; 0,2496]	0,0174	257	p < 0,05
Spearman	ρ = 0,1505	0,0154	—	—	—	p < 0,05

Fonte: Elaborado pelo autor (2026).

Limitações Metodológicas e Recomendações para Refinamento

O processo de validação evidenciou aspectos passíveis de refinamento em futuras iterações. Quanto à validação cruzada, ela não foi adotada nesta fase exploratória, pois o objetivo principal consistiu em testar a viabilidade da abordagem em um subconjunto manualmente rotulado (n=65). Recomenda-se que versões futuras incorporem validação cruzada com 5 ou 10 *folds*. Quanto ao intervalo de confiança do Kappa, embora obtido por *bootstrap* (IC 95%: 0,2687–0,5949), estudos futuros podem ampliar o número de replicações. Quanto à comparação com baseline, não foi realizada formalmente, embora um classificador ingênuo que sempre predissesse TRL 3 alcançaria acurácia de 46%, inferior aos 67,69% obtidos. Quanto à reprodutibilidade, os parâmetros do *Random Forest* foram definidos com *ntree* = 100 e *min_n* = 10.

Essas reflexões integram o processo de amadurecimento do modelo, confirmando que a aplicação de IA em prospecção tecnológica é promissora e exige rigor metodológico progressivo, um movimento análogo à própria escala TRL aplicada à pesquisa.

Os resultados indicam que o modelo é válido para fins de prospecção tecnológica exploratória no âmbito do PPGBIOTEC/UFAM, especialmente na identificação de pesquisas em estágio intermediário (TRL 3–4). A classificação de TRL 5 deve ser tratada com cautela devido ao reduzido número de amostras (n=9). A

literatura recomenda que matriz de confusão e Kappa sejam reportados em conjunto (Banerjee *et al.*, 1999), o que foi seguido nesta análise.

A síntese analítica encontra-se no Quadro 13, que confirma a validade do método para prospecção exploratória.

Quadro 13 – Síntese Analítica dos Resultados

Aspecto	Conclusão
Acurácia final	67,69% - excelente para classificação de textos acadêmicos
Kappa final	0,4206 (IC 95%: 0,2687–0,5949) - concordância moderada
TRL 3 (Médio)	100% de acerto - altamente confiável
TRL 1 (Baixo)	50% de acerto - aceitável
TRL 5 (Alto)	11% - limitado por n amostral
Validade	Válido para prospecção exploratória*

Fonte: Elaborado pelo autor (2026).

Na fase inicial desta pesquisa, técnicas de NLP foram empregadas para identificar o potencial de patenteamento das teses e dissertações do PPGBIOTEC com base nos códigos CIP do conjunto Biotec-BR. Em complemento a essa análise, o método *Random Forest* com features baseadas em *keywords* foi utilizado para classificar os TRL dos mesmos documentos. Esta segunda fase de aprofundamento metodológico de validação incluiu refinamento iterativo da rotulagem, calibração sistemática de pesos, análise evolutiva da matriz de confusão e do coeficiente Kappa, bem como o uso de *bootstrap* para estimação de intervalos de confiança. Esse conjunto de procedimentos confere robustez às evidências apresentadas e estabelece uma base sólida para futuros aprimoramentos.

CAPÍTULO 4 - PREDIÇÃO DE ÁREAS PRIORITÁRIAS PARA BIOPROSPECÇÃO NA AMAZÔNIA: IDENTIFICAÇÃO DE *HOTSPOTS* UTILIZANDO *MACHINE LEARNING*¹¹

O presente capítulo apresenta uma versão reformatada do artigo intitulado “Otimização da Bioprospecção de Ativos Biológicos por Inteligência Artificial”, que constitui o texto-base das seções 4.1, 4.2, 4.3, 4.4 e 4.5. Diferentemente do que foi realizado nos Capítulos 2 e 3, optou-se por não incluir aqui uma seção complementar de aprofundamento metodológico. Essa decisão decorre da natureza exploratória do modelo preditivo desenvolvido, cujo desempenho ($R^2 = 0,0627$) indicou baixo poder explicativo diante da amostra disponível. Como discutido na Seção 4.5, tais limitações estão diretamente relacionadas ao número restrito de observações ($n = 106$) e à complexidade inerente às relações entre variáveis ambientais e atividade biológica.

Nesse contexto, a incorporação de métricas como matriz de confusão ou coeficiente Kappa não se aplica ao modelo apresentado, uma vez que se trata de um problema de regressão, predição de atividade biológica, e não de classificação. A validação foi conduzida por meio das métricas RMSE (*Root Mean Square Error*), MAE (*Mean Absolute Error*) e R^2 , conforme reportado na Seção 4.4, em consonância com as recomendações da literatura para modelos preditivos contínuos (Chai; Draxler, 2014; Willmott; Matsuura, 2005). A decisão de não estender a análise com métodos adicionais decorre do reconhecimento de que o arcabouço metodológico aqui apresentado constitui, sobretudo, uma *proof of concept*, cujo valor reside na demonstração de viabilidade e não na otimização paramétrica exaustiva.

Assim, preserva-se a integridade do artigo publicado como unidade autônoma, sem prejuízo da coerência metodológica da tese. As limitações do modelo são explicitamente discutidas e contextualizadas no próprio capítulo, garantindo transparência e consistência com o *Framework* Sequencial adotado ao longo do trabalho.

¹¹ Este capítulo integra a etapa prescritiva do *Framework* Sequencial apresentado na seção “4.2 *Framework* Sequencial de Inteligência de Dados para Prospeção Tecnológica”, sucedendo as análises de eficiência institucional (Capítulo 2) e de potencialidade tecnológica (Capítulo 3). O presente Capítulo é uma versão reformatada do artigo intitulado “Otimização da Bioprospecção de Ativos Biológicos por Inteligência Artificial”, publicado na Revista RGE Interdisciplinar, Vol. 17, N. 4, p. 1–18, 2026, DOI: 10.56238/revgeov17n4-029, (KUME; COSTA, 2026). O texto base das seções 4.2, 4.3 e 4.4 é reproduzido integralmente, conforme normas de autocitação e tese-artigo.

4.1. INTRODUÇÃO: O CONTEXTO DA BIOPROSPECÇÃO E A URGÊNCIA DA INOVAÇÃO PREDITIVA NA AMAZÔNIA

A diversidade de vida na Terra é um de seus mais valiosos patrimônios, especialmente em regiões tropicais como a Amazônia. Ainda assim, a dimensão extensa e a natureza intrincada dos dados nas áreas de biologia, química e ecologia fazem com que sua análise constitua um desafio científico de complexidade crescente. A procura por compostos com atividade biológica ou a identificação de elementos que podem ser utilizados em medicamentos, na agricultura ou na produção industrial vem passando por uma transição paradigmática, impulsionada pelo progresso da inteligência artificial (IA). Enquanto os métodos tradicionais de investigação exigem muito tempo e recursos financeiros, a IA oferece a possibilidade de converter dados em informações úteis na seleção de componentes que são verdadeiramente valiosos.

Neste contexto, emerge a proposição de investigar: investigar os mecanismos pelos quais os algoritmos de *Machine Learning* conseguem aprimorar a escolha e a priorização das áreas de prospecção de ativos biológicos. Essa abordagem acelera a identificação de ativos biológicos de maneira eficiente, economicamente viável e sustentável.

Na bioeconomia, a velocidade de uma descoberta configura-se como um determinante crítico da competitividade setorial, e os métodos convencionais de triagem biológica têm uma taxa de sucesso comercial bastante baixa em comparação com os altos investimentos que demandam. Por outro lado, técnicas de ML, como *Random Forest* e redes neurais, permitem detectar padrões ocultos em bancos de dados complexos, identificar variáveis com maior poder preditivo e gerar modelos preditivos aplicáveis diretamente na triagem *in silico* de ativos. A junção da bioprospecção com a inteligência artificial transforma a maneira de procurar novos ativos biológicos, ao mesmo tempo, em que diminui a necessidade de abundância de recursos.

Este trabalho apresenta um modelo computacional prototipado em R e Shiny, componentes da plataforma *Open Science*, concebido como um *mockup* funcional para validar a metodologia proposta. A iniciativa não se limita à fundamentação teórica ou à revisão bibliográfica, mas busca uma abordagem prática que permita testar e indicar a aplicabilidade do método. Esse protótipo funcional viabiliza experimentar

conceitos e a análise de sua viabilidade, explorando as condições ambientais mais favoráveis à atividade biológica de compostos e refinando critérios estratégicos, embora não se tenha o propósito de desenvolver um *software* completo.

A metodologia adotada enfatiza aplicar testes de validação, essenciais para garantir a robustez dos modelos desenvolvidos, além da identificação de outliers, que podem comprometer interpretações e conclusões. A integração de atributos físico-químicos, aplicações conhecidas e dados climáticos organizados em camadas espaciais não só permite treinar modelos de classificação, mas também minimizar vieses e aprimorar a confiabilidade das decisões estratégicas de coleta, análise e investimento. Portanto, o objetivo precípuo não consiste em entregar um sistema operacional, mas sim estabelecer um alicerce conceitual que garanta a consistência e pertinência do modelo sugerido nos padrões científicos.

Diante da urgência climática, crise sanitária global e da busca por alternativas sustentáveis aos recursos sintéticos, integrar ciência de dados à bioprospecção passou a ser de uma alternativa facultativa em um imperativo estratégico.

4.2. FUNDAMENTOS CONCEITUAIS: BIODIVERSIDADE, INTELIGÊNCIA ARTIFICIAL E MODELAGEM PREDITIVA

A fundamentação teórica deste estudo repousa sobre a intersecção entre diversidade biológica, inteligência artificial e modelagem preditiva, enquanto os aprimoramentos tecnológicos têm potencial para mudar os encaminhamentos da bioprospecção de ativos biológicos. É a intersecção entre a ciência de dados, geoinformação e bioeconomia que aqui estabelece uma abordagem do referencial teórico, fornecendo suporte metodológico para a construção de modelos mais eficientes, mais sustentados e estrategicamente orientados para a valorização da diversidade biológica.

Biodiversidade e ativos biológicos na Amazônia

Nesta seção do estudo, analisa-se a contribuição da inteligência artificial para acelerar e refinar a descoberta de compostos com potencial bioativo, ao reunir informações ambientais, químicas e biológicas, e integrá-las com técnicas avançadas

de predição, ML e dados georreferenciados, obtém-se uma abordagem mais estratégica e eficiente na exploração da biodiversidade.

Dados geoespaciais para modelagem ecológica

A exploração da rica biodiversidade da Amazônia, seus materiais sustentáveis e metabólitos secundários com usos farmacêuticos, cosméticos e agrícolas, requer métodos de triagem e previsão mais eficazes. *Shapefiles* são, então, constituídos por elementos vetoriais, que podem ser pontos, linhas ou polígonos, delimitando fronteiras, áreas de distribuição e registros de ocorrência. Os *rasters*, diferentemente, estruturam os dados em formatos de grade, nos quais cada ponto retém informações com referência geográfica, representando variações do meio ambiente, a exemplo de temperatura, altitude ou o tipo de vegetação existente (Santos, 2024). Tais formatos, quando combinados, potencializam a análise, permitindo a utilização de dados espaciais em modelos ecológicos, a previsão de áreas de ocorrência de espécies e o estudo dos impactos ambientais de modo mais rigoroso.

Modelagem preditiva com IA

Segundo Prajapati *et al.* (2025), tem se mostrado uma estratégia eficaz, a classificação de áreas de interesse por algoritmos supervisionados como o *Random Forest*, com base na probabilidade de ocorrência de determinados recursos. O uso de algoritmos supervisionados, principalmente *Random Forest*, destaca-se na classificação de áreas favoráveis à ocorrência de ativos específicos, conforme demonstram estudos recentes.

Implantação em R Shiny

Com o uso de pacotes como Shiny (Sievert, 2020), no ambiente integrado de desenvolvimento da linguagem R, RStudio, é possível desenvolver aplicações web interativas, para as análises preditivas, utilizando técnicas de ML como o *Random Forest*, que implementa um conjunto de árvores de decisão e tem sido amplamente utilizado em estudos de ML (Mishra *et al.*, 2024). A combinação dessas ferramentas

permite gerar visualizações como mapas de probabilidade, mapas de calor e cenários de coleta baseados em dados reais, tornando o processo analítico mais acessível e favorecendo a tomada de decisões por pesquisadores e gestores. Além disso, amplia a eficiência e a aplicabilidade das análises ambientais e biológicas.

Dados utilizados

Para os modelos preditivos poderem funcionar, são incorporadas variáveis ambientais, químicas e de biodiversidade, tanto de fontes públicas quanto de dados simulados, provenientes do Sistema de Informação sobre a Biodiversidade Brasileira (SiBBr). É imprescindível a integração dos modelos ambientais e biológicos com informações vetoriais espaciais, geográficas e climáticas para que se possa realizar uma modelagem geoespacial minuciosa e precisa, o que permitirá criar mapas preditivos e planejar de maneira mais direcionada as estratégias de bioprospecção em campo.

Aplicações de IA na prospecção de recursos naturais e exploração geoespacial

Análises que envolvem grandes volumes de dados, imagens e sinais geofísicos, em prospecção de áreas de interesse, beneficiam-se da utilização de inteligência artificial. Áreas de difícil acesso tiveram a identificação de sítios possibilitada pela utilização de imagens de satélite, sensores remotos e algoritmos de detecção em investigações arqueológicas sobre os geoglifos da Costa de Nasca e as construções do Vale de Ica (Karamitrou *et al.*, 2022; Küçükdemirci; Sarris, 2022).

No setor mineral, modelos como redes neurais, árvores de decisão e florestas aleatórias são aplicados ao exame de dados geológicos, geofísicos e geoquímicos, visando a identificar padrões associados à mineralização (Hasan; Shafiq; Khatun, 2023).. Outras abordagens incluem mineração de texto, modelagem conceitual e visualizações baseadas em grafos de conhecimento, que contribuem para o refinamento do processo de descoberta (Li; Chen; Liu, 2022). Técnicas como GNNWLR, GWLR e GWRF também aparecem em estudos recentes, sobretudo pela capacidade de captar variações espaciais e relações não lineares (Wang *et al.*, 2024).

Na indústria de petróleo e gás, métodos supervisionados como *Support Vector Machines* e *Random Forest* são empregados na classificação de regiões geológicas com potencial para exploração, contribuindo para decisões mais precisas (Choubey; Karmakar, 2020).

Em estudos hidrológicos realizados na bacia de Fincha, na Etiópia, redes neurais e métodos hierárquicos analíticos foram empregados para identificar zonas de potencialidade de aquíferos. A integração entre IA e sistemas de informação geográfica tem fortalecido análises espaciais e apoiado a gestão de recursos hídricos (Tamiru; Wagari; Tadese, 2022).

No campo biológico, iniciativas como a da Basecamp Research utilizam algoritmos de ML para catalogar biodiversidade, prever funções proteicas e identificar novas moléculas, com foco em práticas éticas e parcerias internacionais (Johnson, 2025).

Aspectos operacionais, econômicos e de gestão

Os aspectos práticos da inteligência artificial na bioprospecção, nesta seção, são os impactos operacionais, as previsões financeiras e as orientações de gestão que tornam a operação mais eficiente, menos custosa e com decisões mais assertivas e sustentáveis.

Cenários de redução de prospecção com IA vs. Tradicional

Uma vantagem significativa em relação às abordagens tradicionais, para otimizar a celeridade e reduzir os custos na identificação de novos ativos biológicos, se dá pela inteligência artificial (Alloun; Calvio, 2024). Ela ainda possibilita o monitoramento em tempo real das variáveis durante a coleta, elevando tanto a produção quanto o padrão (Pomeroy; Grilc; Likozar, 2022), e estima uma diminuição nos custos operacionais de até 30% (Chan, W. K. C. *et al.*, 2019).

A estimativa de demandas e estoque, juntamente ao setor logístico, a inteligência artificial preditiva é utilizada para planejar suprimentos e gerenciar etapas críticas da cadeia logística e de suprimentos (Joel *et al.*, 2024). A adoção destas tecnologias requer um investimento de capital inicial e formação profissional das

equipes. Em contextos de pouca disponibilidade de dados ou quando procedimentos simples atendem ao objetivo, permanece a vigência das metodologias convencionais de triagem.

Questões operacionais

Na logística operacional e no planejamento de campanhas de prospecção, a IA melhora a coordenação entre equipes e otimiza a logística com apoio de análises preditivas e processamento, em tempo real Danach, Dirani e Rkein (2024). Segundo Rane, Choudhary e Rane (2024), dados históricos analisados por IA são empregados na seleção de áreas para as campanhas. O material orienta a distribuição dos recursos e o planejamento das saídas de campo. Os modelos preditivos podem simular diversos cenários, com a possibilidade de propiciar uma estimativa orçamentária com maior acurácia. A análise preditiva também melhora a previsão de demanda, reduzindo os riscos de excesso ou falta de estoque, conforme apontado por Joel *et al.* (2024).

Aspectos econômico-gestores

A bioprospecção de ativos biológicos, direcionada por IA, otimiza a gestão econômica, em especial na avaliação de ROI, economia financeira, utilização de recursos e administração de riscos.

O ML possibilita observar através da realocação de portfólios em tempo real, um crescimento em torno de 12,5% no retorno sobre o investimento (ROI) (Ambuli *et al.*, 2024; Mediansyah; Septian; Zikry, 2024). A geração de recursos extras e redução de riscos, proporcionada pela análise de grandes volumes de dados feita com apoio de IA torna as estratégias mais rápidas e adaptáveis às mudanças do mercado (Zhu; Hu; Luo, 2024). Na gerência de riscos, sem dúvida, os instrumentos como Redes Neurais Artificiais e Máquinas de Vetores de Suporte aproveitam dados passados para entender as tendências e prever movimentos (Faheem; Aslam; Kakolu, 2022).

Esses modelos preditivos possibilitam que as empresas ajam antes que ocorra um problema, reduzindo riscos oriundos da incerteza do mercado (Gupta; Sharma; Abubakar, 2018).

Impactos na gestão

Além dos aspectos econômicos, os impactos gerenciais também merecem destaque. Segundo (Boardman *et al.*, 2018), modelagens preditivas, mesmo com limitações na precisão dos modelos, são capazes de reduções expressivas de custos operacionais em comparação com métodos empíricos convencionais, técnicas como *Random Forest* possibilitam a economia de recursos quando aplicadas com critério (Giannini *et al.*, 2012).

4.3. ESTRATÉGIA METODOLÓGICA: FONTES, COLETA E CONSTRUÇÃO DO MODELO PREDITIVO

A abordagem utilizada neste trabalho une métodos de modelagem preditiva, fusão de dados ambientais e recursos computacionais para validar a proposta apresentada. O principal objetivo é criar um protótipo funcional que possa replicar cenários e testar a viabilidade da estratégia proposta.

Abordagem Técnica e Desenvolvimento de Modelo

A seguir, os processos técnicos que serviram para moldar o modelo preditivo são explicados, desde a escolha das variáveis e o treinamento com os algoritmos de ML até a adição de dados geoespaciais, para validação da proposta e a simulação dos cenários de prospecção com maior acurácia.

Fontes de Dados e Inserção no Sistema

Os dados biológicos foram extraídos das bases de dados do SiBBr, com registros de espécies e seus atributos. Foram complementados com uma coluna que atribui a cada amostra um nível de interesse, de 1 a 10, com uma casa decimal, com base na atividade biológica e no potencial do ativo biológico identificado. Esse nível orienta a busca pelas áreas prioritárias.

Além disso, os *rasters* climáticos e ambientais, obtidos do Banco de Dados de Variáveis Ambientais para Modelagem de Espécies (AMBDATA) e *shapefiles* do

Ministério do Meio Ambiente (MMA), foram processados utilizando o pacote *raster* no R. Passaram por etapas de filtragem, padronização e normalização, procedimentos que garantem a compatibilidade dos dados com o algoritmo preditivo, permitindo uma modelagem precisa e confiável.

Modelagem Preditiva

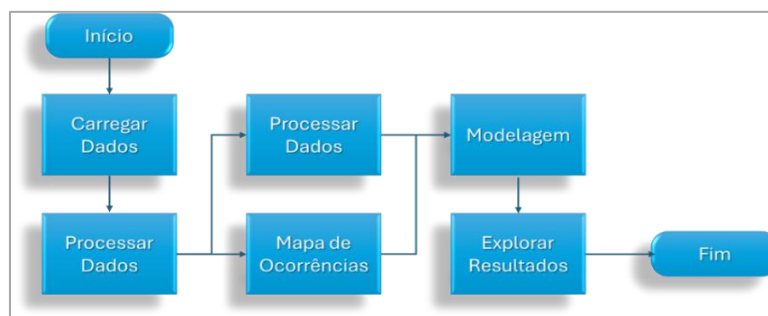
Foram utilizados modelos de *Random Forest*, que foram alimentados com dados ambientais, como precipitação, pressão de vapor, temperatura e uso do solo, além de técnicas para aumentar a precisão das previsões. A solidez do modelo foi testada utilizando um esquema de validação cruzada repetida (*Repeated K-Fold Cross-Validation*), onde os dados de treinamento são divididos em cinco subconjuntos (*folds*), de modo que o modelo seja testado em diversas amostras, repetindo esse procedimento três vezes para minimizar a variabilidade e melhorar a confiabilidade dos resultados (Wong; Yeh, 2020).

A literatura recomenda que, em cenários de baixa explicabilidade global e amostras reduzidas, observações cujo erro absoluto seja substancialmente inferior ao erro médio podem ser classificadas como previsões de alta confiabilidade local. Abordagem essa, consistente com recomendações metodológicas de avaliação de modelos que utilizam RMSE e MAE como referências diretas para definir limites de erro aceitáveis (Chai; Draxler, 2014; Kuhn; Johnson, 2016; Willmott; Matsuura, 2005).

Implementação em R Shiny

Mesmo sendo somente um *mockup*, ele foi estruturado com interface amigável, contendo elementos de visualização de mapas e identificando as regiões com maior potencial para a presença de ativos biológicos, contemplando comandos para exportação de resultados (Jia *et al.*, 2022).

Figura 1 - Diagrama de fluxo do Modelo



Fonte: Elaborado pelos Autores (2026).

Aspectos Operacionais e Estimativas de Cenários

Esta seção apresenta os componentes operacionais envolvidos na aplicação da inteligência artificial, abordando estimativas de cenários, redução de custos e otimização de recursos, com foco na eficiência estratégica das campanhas de prospecção de ativos biológicos.

Definição de Cenários

A inteligência artificial pode reunir dados operacionais e geológicos para criar cenários preditivos, apoiar a administração eficiente de recursos e orientar decisões em tempo real, resultando em eficiência e redução de custos operacionais (Gupta; Sharma; Abubakar, 2018; Monigha, 2024). A exemplo desse potencial, temos o RAP-Optimizer (otimizador de custos inteligente), que pode ter alcançado uma redução de 45% nos custos de servidores (Sathupadi *et al.*, 2024) e de sistemas que, no tempo real, disponibilizam a estratégia a ser utilizada em função das variações de preços dos mercados (Verma *et al.*, 2024). Comparações entre o custo atual e o custo gerado com IA indicam ser possível ter uma economia de até 40% no custo por mês no ambiente de nuvem (Sathupadi *et al.*, 2024) e a redução do custo por barril de petróleo (Gupta; Sharma; Abubakar, 2018).

Avaliação de Risco e Gerenciamento

Dados históricos bem estruturados são a base. Dessa forma modelos preditivos de IA conseguem administrar riscos e alocar recursos com eficiência, da mesma forma que ferramentas de análise preditiva detectam ruídos, mapeiam gargalos e consideram variáveis externas. Já sistemas como o ARON (*Adaptive Resource Optimization Network*) se adaptam, ajustando a estratégia conforme as flutuações e dinâmicas do mercado. O resultado pode chegar a 40% de ganho de eficiência (Verma *et al.*, 2024).

Em situações voláteis e de falta de recursos, a adaptabilidade, oferecida pelos modelos preditivos, torna-se ainda mais significativa, já que permite que as decisões sejam rapidamente ajustadas frente a imprevistos, reduzindo a exposição a riscos que aparecem precisamente nessas situações incertas (Zülfikaroğlu, 2024).

É importante que esses impasses sejam resolvidos cuidadosamente, para que as vantagens prometidas pela tecnologia se manifestem (Samarah *et al.*, 2024; Yesufu; Alajlani, 2025).

Estrutura para a Comparação entre Modelos

Visando comparar a abordagem tradicional de busca biológica com a abordagem proposta baseada em *Random Forest*, foi elaborado um ábaco cruzado, que considera critérios organizados em três eixos: (1) eficiência preditiva; (2) relevância potencial dos achados; e (3) racionalização de custos e recursos logísticos. Cada critério foi avaliado qualitativamente por uma escala semiquantitativa (ex.: Baixo, Médio, Alto), utilizando correspondências visuais para facilitar sua representação gráfica.

Cabe destacar que o objetivo deste exercício não é realizar uma validação empírica exaustiva do modelo proposto, mas sim indicar sua viabilidade metodológica e argumentativa, conforme práticas já observadas em setores como petróleo e gás, onde a aplicação de IA em simulações controladas tem se mostrado eficaz na projeção de cenários e redução de custos operacionais (Gupta; Sharma; Abubakar, 2018; Sathupadi *et al.*, 2024; Verma *et al.*, 2024).

Adicionalmente, os critérios comparativos utilizados neste estudo foram inspirados em *benchmarks* e padrões recorrentes na literatura técnica, especialmente em aplicações de IA voltadas à otimização logística, estimativa de custos e gestão de recursos, com base em dados históricos e estudos de campo previamente documentados (Boardman *et al.*, 2018; Chan, W. K. C. *et al.*, 2019; Danach; Dirani; Rkein, 2024; Rane; Choudhary; Rane, 2024).

O ábaco comparativo elaborado indica que o modelo baseado em IA apresenta potenciais ganhos de eficiência frente à abordagem tradicional. Com relação às tentativas necessárias, o modelo de IA deve indicar um desempenho estimado superior ao método tradicional apresentando índices inferiores, na quantidade de incursões. Tais resultados refletem padrões relatados na literatura, que apontam que a sensibilidade e a acurácia de modelos *Random Forest* podem ultrapassar 90% em cenários com múltiplas variáveis ambientais (Cutler *et al.*, 2007).

Figura 2- Comparação de Métodos de Prospeção

Prospecção Tradicional	Alto	Empírico	Limitada	Indireto	Aleatória	Baixa	Alto	Meses	Alto
Modelo IA (Random Forest)	Baixo	Baseado em dados	Alta	Explícito	Elevada	Alta	Reduzido	Dias/Se m	Reduzido
Critério	Tentativas até achado relevante	Base de decisão	Abrangência espacial	Direcionamento por hotspots	Potencial de inovação disruptiva	Prospecção em áreas remotas	Custo por achado relevante	Tempo médio até decisão	Custo de oportunidade

Fonte: Elaborado pelos autores (2026).

4.4. DINÂMICA E ANÁLISE PROSPECTIVA: RESULTADOS DO MODELO PREDITIVO E IMPLICAÇÕES PARA A GESTÃO DA INOVAÇÃO

No recorte analisado neste estudo, foram utilizados 106 registros referentes ao composto Xantonas, modelados com 9 variáveis preditivas: altitude, Prec_MesColeta, solo, Srad_MesColeta, Tmax_MesColeta, Tmed_MesColeta, Tmin_MesColeta, Vapr_MesColeta e Wind_MesColeta.

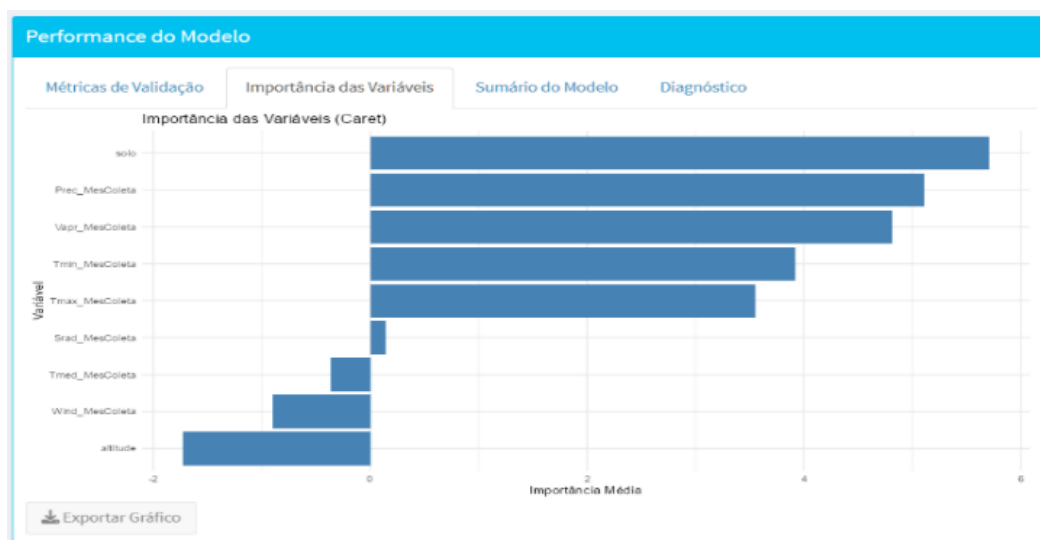
O modelo de *Random Forest* foi ajustado com *ntree* = 500 e *mtry* = 3, utilizando

validação cruzada 5-fold repetida 3 vezes. As métricas de desempenho na validação cruzada indicaram RMSE = 2.4102, MAE = 1.956 e R-quadrado (Coeficiente de Determinação - R^2) = 0.0627, enquanto a variância explicada *Out-of-Bag* (OOB) do modelo final foi de -26,29%. Esses resultados sugerem que, embora o modelo capture parte da estrutura dos dados, ainda apresenta baixo poder explicativo global, reforçando a necessidade de refinamento paramétrico e ampliação da base de dados.

A Figura 3 apresenta a importância relativa das variáveis preditivas no modelo *Random Forest*, evidenciando que o tipo de solo, precipitação e pressão de vapor foram os principais determinantes da atividade prevista.

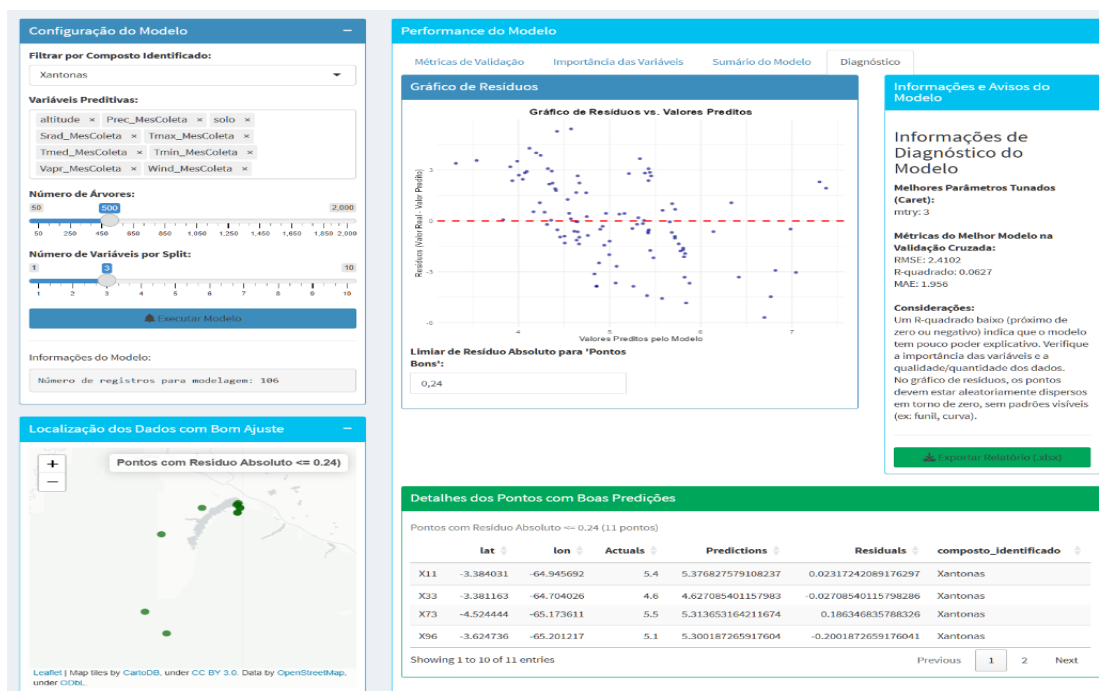
Para fins de diagnóstico, adotou-se um limiar de resíduo absoluto de 0,24, definido de forma proporcional ao desempenho global do modelo. Esse valor corresponde a aproximadamente 10% do RMSE (2,41) ou 12% do MAE (1,96), situando-se claramente abaixo do erro típico das previsões. A escolha desse limiar segue princípios amplamente reconhecidos na literatura de modelagem preditiva e ecológica, segundo os quais resíduos substancialmente inferiores às métricas globais de erro caracterizam regiões de alta confiabilidade local, especialmente em modelos com baixa variância explicada ou amostras reduzidas (Anderson *et al.*, 2011; Elith; Leathwick, 2009; Franklin; Miller, 2009). Sob esse critério conservador, os pontos com $|\text{resíduo}| \leq 0,24$ representam estimativas localmente estáveis, funcionando como ilhas de confiabilidade preditiva úteis para orientar análises exploratórias e sustentar a prova de conceito da etapa de bioprospecção, Figura 4.

Figura 3 - Importância das Variáveis Preditivas



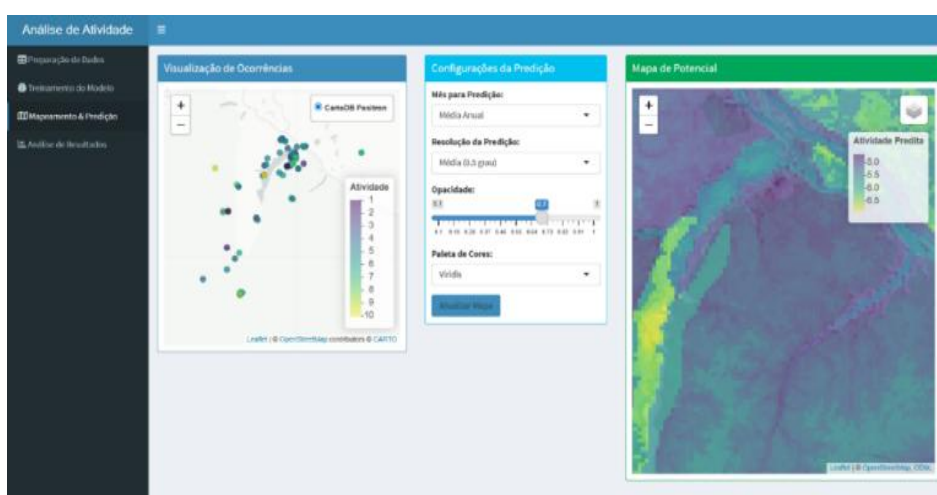
Fonte: Elaborado pelos autores (2026).

Figura 4 - Diagnóstico do Modelo



Fonte: Elaborado pelos autores (2026).

A Figura 5 apresenta o mapa de *hotspots* gerado pelo modelo, destacando as áreas com maior probabilidade de ocorrência de atividade biológica elevada. Observa-se a formação de gradientes espaciais coerentes com os padrões ambientais da região.

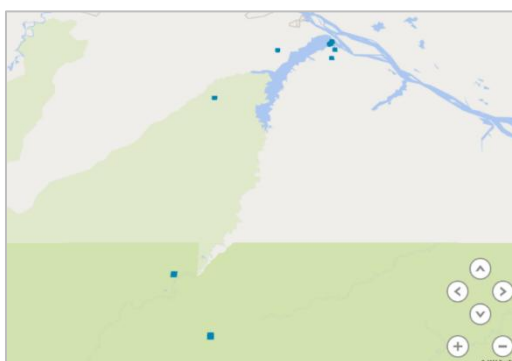
Figura 5 - Mapa de *Hotspots*

Fonte: Elaborado pelos autores (2026).

Sob esse critério conservador, foram identificados os pontos com $|\text{resíduo}| \leq 0,24$, os quais representam predições com erro muito inferior ao erro típico do modelo

e, portanto, constituem candidatos prioritários para prospecção, mesmo diante do baixo número de amostras ($n = 106$) e da variância explicada reduzida. Esses pontos funcionam como ilhas de estabilidade preditiva, úteis para orientar ajustes futuros e para indicar que, apesar das limitações globais, o modelo apresenta potencial preditivo local válido Figura 6.

Figura 6 - Pontos com $|\text{resíduo}| \leq 0,24$



Fonte: Elaborado pelos autores (2026).

As limitações observadas podem estar associadas à quantidade restrita de amostras, que pode não ser suficiente para capturar a complexidade das relações entre variáveis ambientais e a atividade. Além disso, a escolha de somente quatro variáveis ambientais pode limitar a explicação dos padrões observados, uma vez que fatores ecológicos, biológicos ou antrópicos não considerados podem influenciar significativamente os resultados. Para fortalecer o modelo, algumas estratégias podem ser adotadas:

- Expansão do conjunto de dados, tanto em quantidade e diversidade de amostras quanto em incorporação de variáveis adicionais (topográficas, uso e cobertura do solo, proximidade a corpos d'água, infraestrutura humana, variáveis biológicas e sociais).
- Configurar os hiperparâmetros do *Random Forest* para investigar diversas combinações de *mtry* e *ntree*, identificando a configuração mais apropriada.
- Testes com outros algoritmos de aprendizado, como modelos aditivos generalizados (GAMs), máquinas de vetores de suporte (SVMs) e abordagens baseadas em *boosting*, para avaliar métodos alternativos de modelagem preditiva, ajustes que podem contribuir significativamente para a melhoria do desempenho

preditivo, possibilitando uma maior precisão e confiabilidade nas estimativas da atividade biológica.

4.5. CONSIDERAÇÕES FINAIS E A TRANSIÇÃO PARA O MODELO INTEGRADOR

A bioprospecção encontra na integração da inteligência artificial com dados espaciais, uma abordagem ajustada e conveniente para a busca de ativos biológicos relevantes. A metodologia racionaliza atividades que, na forma tradicional, exigiriam longos períodos para planejamento, altos custos e muitos recursos logísticos, auxiliando a identificação de áreas de maior potencial, reduzindo custos, encurtando prazos de análises e expandindo o alcance de coletas, tudo isso a favor de uma exploração mais efetivada da biodiversidade.

Por sua natureza interativa e escalável, a solução facilita o acesso a análises avançadas e organiza fluxos de trabalho que fortalecem a tomada de decisão em contextos de prospecção. Além disso, a metodologia empregada minimiza vieses na seleção de áreas prioritárias, promovendo avaliações mais objetivas e fundamentadas em padrões identificados pelo modelo preditivo, aumentando a confiabilidade dos resultados.

Diferentemente da abordagem tradicional, baseada predominantemente na experiência empírica de campo, o modelo orienta a prospecção para *hotspots* de biodiversidade inferidos a partir de variáveis ambientais como precipitação, altitude e cobertura vegetal. Embora a relevância absoluta de cada achado ainda não possa ser quantificada diretamente, a priorização de áreas com maior probabilidade de conter espécies endêmicas ou subexploradas se alinha aos objetivos estratégicos da bioeconomia.

Mesmo sem dados absolutos de custos, o ábaco comparativo indica que a racionalização das expedições proporcionada pelo modelo tem potencial para reduzir significativamente o custo por descoberta relevante e diminuir o tempo entre o planejamento e a execução de novas coletas. O custo de oportunidade associado a equipes especializadas, deslocamentos logísticos e tentativas malsucedidas pode, assim, ser substancialmente mitigado.

A pesquisa realizada não apenas confirma a viabilidade e a utilidade da abordagem proposta, como também estabelece uma base metodológica sólida para

estudos futuros. Os resultados obtidos reforçam o potencial da inteligência artificial como ferramenta estratégica para ampliar a eficiência, a precisão e o alcance das iniciativas de bioprospecção voltadas à bioeconomia.

4.6. ANÁLISE COMPLEMENTAR: NATUREZA EXPLORATÓRIA E CRITÉRIOS DE CONFIABILIDADE LOCAL

O modelo preditivo apresentado neste capítulo foi concebido como prova de conceito metodológica, e não como modelo preditivo final. Seu desempenho global, R^2 de 0,0627 e variância explicada Out-of-Bag de -26,29%, indica baixo poder explicativo, o que é coerente com a natureza exploratória do exercício e com as limitações da amostra disponível ($n = 106$). A utilidade do modelo não reside, portanto, na capacidade de generalização para toda a área de estudo, mas na identificação de pontos de alta confiabilidade local, que podem orientar investigações futuras.

A definição desses pontos baseou-se em um limiar de resíduo absoluto de 0,24. Esse valor não foi escolhido arbitrariamente: ele corresponde a aproximadamente 10% do RMSE (2,41) e 12% do MAE (1,96) do modelo, situando-se claramente abaixo do erro típico das predições. Sob esse critério conservador, os pontos com $|\text{resíduo}| \leq 0,24$ representam estimativas localmente estáveis, ainda que o modelo, como um todo, tenha desempenho preditivo global limitado.

É importante esclarecer que o limiar de 0,24 não constitui uma validação empírica dos *hotspots* identificados, mas um critério de triagem exploratória. A validação empírica exaustiva demandaria ampliação da amostra, coleta de dados de campo e verificação independente da atividade biológica nos pontos identificados — o que está além do escopo desta tese e constitui recomendação para trabalhos futuros. O que o modelo entrega, portanto, não é uma previsão definitiva, mas uma priorização informada de áreas para investigação futura, o que é coerente com a natureza exploratória do capítulo e com o papel do método no Framework Sequencial.

A distinção entre prova de conceito e modelo preditivo final é fundamental para a leitura correta deste capítulo. A prova de conceito demonstra que a abordagem metodológica é viável e internamente consistente; o modelo preditivo final exigiria validação empírica, ampliação da amostra e refinamento dos hiperparâmetros. Essas etapas são explicitamente recomendadas na Seção 4.5 como trabalhos futuros.

CAPÍTULO 5 - *FRAMEWORK* SEQUENCIAL INTEGRADOR: DASHBOARD INTERATIVO E COOPERAÇÃO ABERTA

Os capítulos anteriores apresentaram, de forma sequencial, diferentes dimensões do ecossistema de inovação da UFAM e, em particular, aspectos relacionados à dinâmica tecnológica do PPGBIOTEC. O Capítulo 1 contextualizou a distribuição regional de patentes em biotecnologia; o Capítulo 2 analisou o desempenho da UFAM em comparação com outras instituições públicas; o Capítulo 3 identificou indícios de potencial inventivo ainda não formalizado; e o Capítulo 4 explorou possibilidades analíticas associadas à maturidade tecnológica e à bioprospecção na Amazônia. Em conjunto, esses resultados oferecem uma visão integrada de elementos que influenciam os processos de prospecção tecnológica e de gestão do conhecimento na instituição.

Este capítulo, por sua vez, integra e operacionaliza essas descobertas por meio de um *Framework* Sequencial que culmina em um *Dashboard* Interativo. Embora não represente a contribuição central da tese, o *framework* configura-se como um arranjo metodológico original, não identificado na literatura consultada, para organizar, em fluxo contínuo, análises de desempenho institucional, prospecção tecnológica e modelagem preditiva. Seus elementos podem ser replicados e adaptados por outras Instituições Científicas, Tecnológicas e de Inovação (ICTs) que desejem organizar ou representar análises de natureza semelhante. O *Dashboard* Interativo, concebido como modelo conceitual, sintetiza visualmente os resultados obtidos ao longo do processo analítico.

5.1. O *FRAMEWORK* SEQUENCIAL COMO CONTRIBUIÇÃO METODOLÓGICA

O *Framework* Sequencial proposto organiza-se em quatro etapas, cada uma correspondente a um dos capítulos analíticos da tese:

Quadro 5 - Etapas do *Framework* Sequencial

Etapas	Método	Produto	Capítulo
1	Mapeamento de patentes (Biotec-BR)	Panorama setorial e posicionamento da UFAM	Cap. 1
2	Análise Envolvória de Dados (DEA)	Eficiência relativa da UFAM e <i>benchmarking</i>	Cap. 2

Etapa	Método	Produto	Capítulo
3	NLP + <i>Random Forest</i>	Potencial de patenteabilidade e classificação TRL	Cap. 3
4	ML preditivo (<i>Random Forest</i> + dados geoespaciais)	Identificação de <i>hotspots</i> para bioprospecção	Cap. 4

Fonte: Elaborado pelo autor (2026).

É necessário esclarecer a natureza deste *Framework*. Não é um produto tecnológico, *software* ou ferramenta de gestão a ser adotada, mas sim um método. Refere-se a uma sequência determinada de ações e análises concebidas para organizar o processo de pesquisa da tese. Sua função básica é compilar evidências, como a eficiência relativa da UFAM e o potencial inventivo estimado no PPGBIOTEC, possibilitando uma interpretação sistemática dessas informações, sem a intenção de direcionar decisões institucionais. Essa distinção define o âmbito estritamente acadêmico e metodológico da pesquisa.

A combinação dessas fases proporciona uma perspectiva coesa do processo de inovação em uma ICT, desde o diagnóstico de desempenho até a identificação de ativos tecnológicos. Apesar de ter sido criado para estruturar o percurso analítico desta tese, o *framework* pode ser utilizado em outros contextos de pesquisa, desde que os conjuntos de dados e os parâmetros utilizados sejam ajustados.

A replicabilidade do Framework Sequencial em outras ICTs depende de três condições: (i) adaptação do conjunto de CIPs ao setor de atuação da instituição (o Biotec-BR é específico para biotecnologia, mas o INPI disponibiliza conjuntos equivalentes para outros setores); (ii) disponibilidade de dados comparáveis para a etapa DEA (insumos e outputs institucionais); e (iii) corpus textual acessível para a etapa NLP/ML (teses, dissertações ou relatórios técnicos), coletado por meio de protocolo OAI-PMH a partir do repositório institucional da ICT. O método é, portanto, replicável como recomendação, e não como protocolo exaustivo: exige calibração setorial e institucional, e eventuais adaptações dependerão das especificidades de cada ICT.

5.2. ARQUITETURA DO *DASHBOARD* INTERATIVO

O *Dashboard* Interativo foi concebido como uma materialização ilustrativa do *Framework* Sequencial. Desenvolvido em R/Shiny, ambiente de código aberto

alinhado aos princípios da Ciência Aberta, o *dashboard* integra quatro camadas de informação:

- I. Camada de eficiência institucional (DEA): escores de eficiência da UFAM em comparação com outras IES, com destaque para *benchmarks* e gargalos identificados.
- II. Camada de potencial de patenteamento (CIP): visualização das 55 pesquisas com aderência aos códigos Biotec-BR, distribuídas por ano e área tecnológica.
- III. Camada de maturidade tecnológica (TRL): classificação das 259 teses e dissertações por nível TRL, com filtros por confiança de predição e faixa de desenvolvimento.
- IV. Camada de prospecção preditiva (ML): mapas de *hotspots* para bioprospecção na Amazônia, indicando áreas prioritárias para coleta e validação de ativos biológicos.

A arquitetura técnica adota o R como linguagem de processamento e visualização, com as bibliotecas *shiny*, *shinydashboard*, *plotly*, *ggplot2* e *DT*. Os dados são armazenados em formato tabular (CSV/Excel) e atualizáveis periodicamente.

5.3. FUNCIONALIDADES E NAVEGAÇÃO

O *Dashboard* permite ao usuário:

- Consultar o portfólio de ativos tecnológicos do PPGBIOTEC, com identificação parcialmente anonimizada para preservar estratégias de proteção intelectual.
- Aplicar filtros por ano, TRL, código CIP, confiança de predição e potencial de patenteamento.
- Visualizar a matriz TRL × CIP, identificando cruzamentos entre maturidade e áreas tecnológicas.
- Explorar mapas de *hotspots* para bioprospecção, com camadas sobrepostas de variáveis ambientais (solo, precipitação, temperatura, pressão de vapor).
- Exportar relatórios em formatos CSV e Excel para análises complementares.

5.4. COOPERAÇÃO ABERTA E TRANSPARÊNCIA

O *Dashboard* foi concebido como um instrumento de Cooperação Aberta, alinhado ao modelo da Hélice Quíntupla, que incorpora academia, indústria, governo,

sociedade civil e ambiente natural. Ao disponibilizar dados estratégicos de modo transparente e acessível, o *dashboard*:

- Contribui para reduzir assimetrias de informação entre o PPGBIOTEC e as estruturas administrativas de inovação da UFAM.
- Aproxima pesquisadores do setor produtivo, ao revelar ativos tecnológicos com potencial de licenciamento.
- Amplia o controle social, permitindo que a sociedade acompanhe o retorno dos investimentos públicos em pesquisa e inovação.
- Estimula parcerias com comunidades tradicionais, ao integrar dados de biodiversidade e conhecimentos associados.

A transparência promovida pelo *Dashboard* não é um fim em si, mas um meio para apoiar a superação da inércia institucional diagnosticada nos Capítulos 1 e 2.

Integração Interdisciplinar necessária para a inovação em biotecnologia

A inovação em biotecnologia exige a convergência de múltiplas competências científicas, tecnológicas e gerenciais. No contexto da UFAM, essa integração pode ser fortalecida, uma vez que diferentes áreas detêm conhecimentos complementares que, quando articulados, ampliam a capacidade institucional de organizar informações, identificar oportunidades e orientar processos de inovação. O *framework* concebido e desenvolvido como método desta tese sinaliza que a cooperação necessária para a inovação não é apenas externa, mas também interna, envolvendo unidades acadêmicas e administrativas da própria instituição.

Essa integração pode ser exemplificada pelas seguintes áreas, mas não se limita a elas:

- I. Biotecnologia, Biologia e Farmácia - produção de conhecimento biológico, identificação de bioativos, validação experimental e caracterização funcional. Essas áreas também fornecem a base conceitual necessária para a rotulação manual utilizada no treinamento dos modelos de *Machine Learning*, especialmente na classificação TRL e na identificação de elementos técnicos relevantes para a análise de patenteabilidade.;
- II. Engenharia de Computação e Engenharia de Dados - aplicação de NLP, ML, classificação TRL, automação analítica e integração de dados ambientais e

institucionais. Essas áreas operacionalizam a rotulação realizada pelas áreas biológicas, transformando-a em *corpus* estruturado para treinamento supervisionado, vetorização, validação e classificação automática da base completa;

- III. Química de Processos - desenvolvimento tecnológico, escalonamento, purificação e avaliação de viabilidade industrial dos ativos identificados;
- IV. Administração e Gestão da Inovação - estratégia, métricas de inovação, avaliação de maturidade tecnológica, tomada de decisão e desenho de políticas institucionais;
- V. PROTEC (NIT da UFAM) - proteção da propriedade intelectual, gestão de ativos tecnológicos e articulação com o setor produtivo.

5.5. LIMITAÇÕES E PRÓXIMOS PASSOS

O *Dashboard* apresentado é um modelo conceitual, não um sistema em produção. Sua implementação plena na UFAM demandaria: (i) integração com sistemas institucionais de gestão da inovação; (ii) validação com usuários reais (PROTEC, coordenadores do PPGBIOTEC, pesquisadores); e (iii) atualização periódica dos dados de patentes e produção acadêmica.

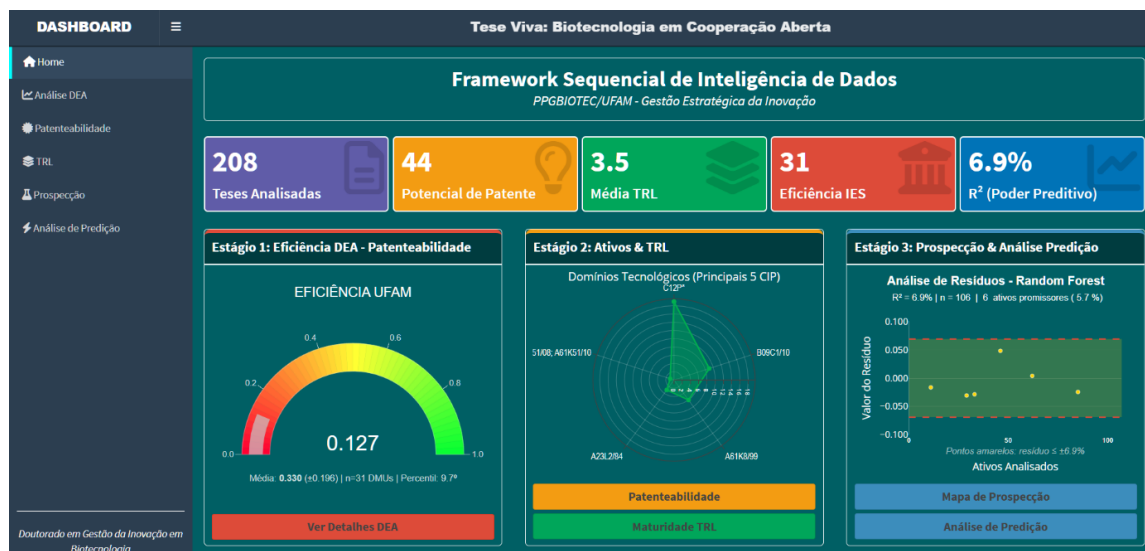
Para trabalhos futuros, recomenda-se: (a) replicar o *Framework* Sequencial em outras ICTs da Amazônia Legal; (b) testar modelos semânticos avançados (BERTimbau, BioBERT) para aprimorar a classificação TRL; (c) acompanhar longitudinalmente os 55 ativos identificados para verificar sua efetiva conversão em patentes ou parcerias.

A tese não tem por objetivo prescrever o caminho institucional de adoção de qualquer método por ela abordado, nem estimar custos, prazos ou fontes de financiamento para essa finalidade. Essas decisões envolvem variáveis institucionais, orçamentárias e políticas que fogem ao escopo desta pesquisa. O que a tese oferece é a demonstração da viabilidade metodológica da abordagem, cabendo às instâncias competentes da UFAM (a PROTEC, o CONSUNI e a Reitoria) a eventual decisão sobre implementação.

Em uma agenda de pesquisa distinta, potencialmente no âmbito do Programa de Pós-Graduação em Propriedade Intelectual e Transferência de Tecnologia para a Inovação (PROFNIT), poder-se-ia explorar, em parceria com a PROTEC, um plano de

adoção institucional, com definição de cronograma, orçamento e governança de dados. Essa é uma agenda que se beneficiaria dos achados aqui apresentados, ainda que não integre os objetivos desta tese.

Figura 5 - *Dashboard* Interativo de Cooperação Aberta



Fonte: Elaborado pelo autor (2025).

4. SÍNTESE E PERSPECTIVAS

4.1. SÍNTESE DOS ACHADOS

No caso da UFAM, o modelo DEA aplicado no Capítulo 2, com base nos dados apresentados no Capítulo 1, sugere que a instituição não se encontra na fronteira de eficiência, situando-se na faixa de Baixa Eficiência (0 a 0,24) da escala normalizada. Esse posicionamento não implica ausência de potencial científico, mas evidencia a distância relativa em relação às instituições eficientes, considerando os insumos e produtos selecionados. Trata-se de uma leitura descritiva que indica a existência de variações na relação entre recursos disponíveis e resultados observados, padrão comum em instituições inseridas em ecossistemas regionais em consolidação.

No âmbito institucional, a análise complementar apresentada na Seção 2.6 contribui para ampliar a compreensão metodológica do modelo DEA ao indicar que, além de indicar o nível relativo de eficiência das IES, o método permite observar informações adicionais por meio da matriz de lambdas e da DMU virtual. Esses elementos apontam combinações de insumos e resultados associadas a diferentes níveis de eficiência de uma instituição, servindo como referência analítica para interpretação dos resultados. Dessa forma, o DEA deixa de ser apenas um instrumento de diagnóstico e assume papel analítico ao organizar possíveis relações entre insumos e resultados e de fortalecimento das políticas internas de inovação. Movimentos em direção à fronteira de eficiência costumam estar associados a mudanças nos indicadores de desempenho utilizados em avaliações externas. No caso específico da UFAM, conforme evidenciado na Seção 2.6, a matriz de lambdas e a DMU virtual podem ser utilizados como referência metodológica para interpretar o posicionamento da instituição no conjunto analisado.

O Capítulo 3 demonstra que o PPGBIOTEC produz conhecimento com aderência tecnológica mensurável: 21,2% das 259 teses e dissertações analisadas apresentam potencial de patenteamento, incluindo 55 ativos classificados com TRL igual ou superior a 3 e confiança acima de 0,6. Apesar disso, 94,6% das pesquisas situam-se nos níveis TRL 1-3, o que sugere que grande parte da produção científica se encontra em fases preliminares do desenvolvimento tecnológico. Um aspecto metodológico observado na análise do coeficiente de correlação de Pearson é a baixa associação entre maturidade tecnológica e potencial de patenteamento. Esse

resultado sugere que elementos inventivos podem aparecer ainda em fases conceituais, antes de validação experimental ou prototipagem, reforçando a importância de métricas de terceira geração para identificar oportunidades tecnológicas precoces e estruturar *pipelines* estratégicos de inovação.

A análise complementar da Seção 3.7 estende esse conjunto de resultados ao avaliar estatisticamente o modelo de classificação TRL, indicando a evolução das métricas ao longo de sucessivos refinamentos na rotulagem, na calibração de pesos e na distribuição amostral. A avaliação por matriz de confusão e coeficiente Kappa permitiu examinar, ao nível analítico mais refinado, os padrões de acerto e erro do classificador, revelando tanto sua robustez na identificação de pesquisas em estágio intermediário quanto os vieses inerentes à linguagem utilizada nos resumos acadêmicos. A ocorrência de confusões específicas sugere que a proximidade semântica, como entre TRL 1 e TRL 3, ou entre TRL 5 e TRL 3, pode influenciar o desempenho do modelo. Ainda assim, o classificador apresentou utilidade relevante como ferramenta de prospecção exploratória, permitindo organizar uma leitura estruturada das tendências tecnológicas do PPGBIOTEC e identificar ativos que podem ser considerados em estudos posteriores.

A identificação dos 55 ativos com potencial de patenteamento que não avançaram para proteção formal indica a existência do que foi denominado vale da morte institucional: a lacuna entre a produção científica e a proteção intelectual. Esse fenômeno tem sido interpretado como a primeira etapa de um ciclo de inovação interrompida. A etapa seguinte, caracterizada como Ortotanásia Patentária, diz respeito à permanência de ativos já protegidos em estado de inatividade institucional. Quando analisados em conjunto, os achados dos Capítulos 2 e 3 sugerem um ciclo fragmentado: em um primeiro momento, a instituição não converte integralmente oportunidades de proteção; em um segundo momento, ativos já protegidos não avançam para processos de transferência. Esses movimentos refletem dinâmicas observadas em ambientes de inovação em desenvolvimento, nos quais a integração entre produção científica e inteligência de dados aplicada à gestão tecnológica ocorre de forma gradual.

A análise conjunta desses achados sugere ainda que esse ciclo fragmentado não é apenas o resultado de práticas gerenciais internas, mas de condicionantes estruturais do Sistema Nacional de Inovação em si. Dados nacionais apontam que a fraca conversão de insumos em produtos tecnológicos é uma constante no Brasil, o

que leva a crer que a atuação dos NITs se foca em questões jurídico-administrativas e no cumprimento formal de normas. Esta pesquisa sugere uma leitura adicional para esse padrão, observando o arcabouço legal de CT&I como um conjunto de *lex imperfecta*: disposições que impõem deveres institucionais, mas não preveem sanções por descumprimento. Essa característica normativa contribui para a persistência da custódia passiva de ativos tecnológicos, reforçando tanto o vale da morte institucional quanto a Ortotanásia Patentária, ao limitar a ativação tecnológica e a transferência efetiva de conhecimento.

Essa análise indica que o desafio observado não se relaciona à capacidade de produzir conhecimento, mas à inexistência de mecanismos sistemáticos de monitoramento, valoração e prospecção ativa do portfólio tecnológico. A mudança desse padrão é frequentemente tratada em estudos sobre inovação e gestão tecnológica, que descrevem a passagem de uma lógica de custódia passiva para modelos que incorporam acompanhamento contínuo e análise sistemática do portfólio. Os achados dos Capítulos 2 e 3 indicam que esse tipo de abordagem, ainda não plenamente incorporado às práticas institucionais da UFAM, pode contribuir para análises mais estruturadas do portfólio tecnológico.

O Capítulo 4, por sua vez, apresenta uma abordagem adicional às análises de NLP e ML do Capítulo 3, descrevendo caminhos analíticos que permitem observar como novas descobertas podem evoluir para estágios mais elevados de TRL. Com o objetivo de explorar o potencial científico associado à biodiversidade amazônica, o capítulo discute a bioprospecção guiada por Inteligência Artificial como abordagem alinhada à concepção da Amazônia como laboratório vivo, entendida aqui como um arcabouço analítico que integra diversidade biológica, conhecimento científico e possibilidades de aplicação tecnológica. Essa abordagem amplia a capacidade de reconhecer ativos biológicos promissores e orienta os esforços de pesquisa para áreas com maior potencial de inovação, consolidando a transição do conhecimento científico para aplicações tecnológicas tangíveis, ao mesmo tempo, em que reduz custos e a necessidade de recursos na etapa de prospecção.

O modelo preditivo apresentado no Capítulo 4 é meramente metodológico, uma prova de conceito. Apesar do baixo poder explicativo geral ($R^2 = 0,0627$), os resultados indicam que o arcabouço é viável para identificar *hotspots* (Cap. 4 Fig. 5). As limitações identificadas sinalizam a necessidade de aumentar a base de dados e aprimorar os hiperparâmetros, orientando assim pesquisas futuras. O ábaco

comparativo (Cap. 4 Fig. 2) indica o potencial teórico de ganho de eficiência, não se tratando de uma constatação empírica deste estudo.

Contudo, o modelo evidencia que estratégias preditivas podem orientar a alocação de recursos na bioprospecção, contribuindo para a redução dos custos operacionais e aumentando a eficiência das pesquisas de campo. No contexto da UFAM, esses resultados podem ser considerados em discussões sobre o uso de abordagens de inteligência artificial em pesquisas aplicadas, particularmente em setores de elevada complexidade territorial, como a Amazônia.

Para fins de síntese estratégica, os principais achados da tese podem ser organizados segundo quatro dimensões, conforme o Quadro abaixo.

Quadro 6 - Síntese dos Achados por Dimensão Estratégica

Dimensão	Achado da Tese	Capítulo
Potencial institucional	PPGBIOTEC apresenta 21,2% de pesquisas com potencial de patenteamento	Cap. 3
Eficiência relativa	UFAM posiciona-se na faixa de Baixa Eficiência da análise DEA	Cap. 2
Oportunidade metodológica	<i>Framework</i> sequencial e <i>Dashboard</i> indicam viabilidade técnica	Cap. 5
Contexto concorrencial	Concentração de patentes no Sul/Sudeste e presença internacional na Amazônia	Caps. 1 e 4

Fonte: Elaborado pelo autor (2026).

O Capítulo 5 integra esses instrumentos em um *Dashboard* Interativo, apresentando o *Framework* Sequencial como contribuição metodológica que pode ser utilizada em diferentes contextos de pesquisa. Assim, a investigação não se limita ao diagnóstico e abre espaço para refletir sobre diferentes formas de interpretar práticas institucionais relacionadas à inovação.

Portanto, esses achados sugerem que a UFAM reúne potencial científico e condições institucionais que podem favorecer análises relacionadas à gestão da inovação. O *Framework* Sequencial, enquanto método de desenvolvimento da tese, permitiu revelar padrões, lacunas e oportunidades que se materializam no *Dashboard* Interativo e na proposta de Cooperação Aberta, instrumentos que organizam informações relevantes para estudos sobre gestão da inovação orientada por evidências.

A utilização contínua desses instrumentos pode resultar em uma *Alive Thesis*: uma produção científica que permanece viva, em evolução e integrada às dinâmicas

reais da instituição, mantendo a produção científica em evolução mesmo após a defesa.

À luz desse conjunto de evidências, os resultados também apontam que a Gestão da Inovação em Biotecnologia desempenha papel relevante à realização da missão científica e da função social das universidades da região. Em um país historicamente organizado em torno de commodities, extrativismo e ciclos econômicos de baixa complexidade, no contexto amazônico, diferentes estudos discutem modelos institucionais que relacionam biodiversidade, conhecimento científico e desenvolvimento tecnológico.

O PPGBIOTEC, ao integrar pesquisa avançada, formação de recursos humanos e vocação regional, desempenha papel específico nesse processo. A consolidação de práticas de gestão da inovação, incluindo prospecção tecnológica, análise de maturidade, inteligência de dados e articulação com estruturas como a PROTEC, pode contribuir para ampliar a participação da UFAM em iniciativas relacionadas à bioeconomia amazônica.

As evidências produzidas nesta tese sugerem que a inovação não emerge apenas do laboratório, mas da existência de mecanismos institucionais capazes de reconhecer valor, reduzir assimetrias de informação e apoiar processos decisórios relacionados à pesquisa e ao desenvolvimento tecnológico. “Nesse sentido, a Gestão da Inovação em Biotecnologia constitui um elemento central nos processos de ativação tecnológica analisados na tese, especialmente no que se refere ao rompimento da lógica de custódia passiva.

4.2. CONTRIBUIÇÕES DA TESE

Esta tese oferece um conjunto integrado de contribuições teóricas, metodológicas, práticas e institucionais:

Quadro 7 - Contribuições

Tipo	Contribuição
Metodológica	Framework Sequencial (DEA → NLP/TRL → ML → <i>Dashboard</i>) para diagnóstico e prospecção em ICTs, replicável em outros contextos.
Teórica	Proposição e fundamentação do conceito de Ortotanásia Patentária ; identificação do efeito das <i>lex imperfecta</i> no Sistema Nacional de

Tipo	Contribuição
	Inovação; indicação empírica da baixa correlação entre TRL e patenteabilidade (Pearson = 0,1318).
Métrica	Aplicação de Métricas de Terceira Geração à avaliação da transferência de tecnologia, incorporando maturidade, potencial inventivo e impacto socioambiental.
Analítica	Análise DEA aplicada à eficiência institucional em biotecnologia, com comparação entre 29 IES públicas e identificação de <i>benchmarks</i> .
Tecnológica	NLP para revelação de potenciais tecnológicos ocultos em teses e dissertações; ML com <i>Random Forest</i> para otimização e redução de custos em prospecção de ativos biológicos em campo.
Ferramental	Implementação em Open Science (R/Shiny), com código aberto, auditável e alinhado aos princípios de transparência e reprodutibilidade.
Prática	Dashboard Interativo como instrumento de apoio à tomada de decisão.
Conceitual	Cooperação Aberta como princípio analítico e organizador da relação entre academia, setor produtivo, governo e comunidades.
Institucional	Conceito de Alive Thesis - uma tese viva, que permanece após a defesa como instrumento contínuo de inovação institucional.

Fonte: Elaborado pelo autor (2026).

O *Framework* Sequencial consolida-se como uma ferramenta metodológica de terceira geração, capaz de antecipar oportunidades, revelar ativos ocultos e orientar decisões estratégicas em ICTs brasileiras.

Contribuição Político-Estratégica

Além das contribuições organizadas no Quadro 6, esta pesquisa também apresenta uma contribuição político-estratégica de um caráter mais amplo. Embora a pesquisa tenha sido realizada na Universidade Federal do Amazonas, no Programa de Pós-Graduação em Biotecnologia e na Pró-Reitoria de Inovação Tecnológica, os desafios identificados são consistentes com aqueles amplamente relatados na política nacional de Ciência, Tecnologia e Inovação, particularmente no que tange à transformação da produção científica em ativos tecnológicos com relevância socioeconômica. A avaliação realizada enfatiza a necessidade de consolidação de práticas estruturadas à gestão da inovação, as quais se mostram essenciais para a implementação e o cumprimento das orientações da Política de Desenvolvimento da Biotecnologia (Decreto 6.041/2007), da Lei de Inovação (Lei 10.973/2004) e do Marco Legal de Ciência, Tecnologia e Inovação (Lei 13.243/2016). A falta de mecanismos eficazes que realmente promovam a transferência de tecnologia representa uma limitação significativa ao progresso científico e tecnológico do país.

É nesse sentido que a área de concentração em Gestão da Inovação do PPGBIOTEC se torna estratégica. Os resultados desta pesquisa indicam que esse campo de estudo ocupa um espaço acadêmico primordial na formação de profissionais que podem intervir nos gargalos mapeados ao longo da investigação, favorecendo a articulação entre produção científica, maturidade tecnológica e políticas de inovação. Seu fortalecimento é consistente com as evidências apresentadas e é um requisito institucional para que as universidades brasileiras se tornem mais eficazes em suas políticas de inovação, em sintonia com o desenvolvimento socioeconômico da nação.

4.3. PERSPECTIVAS FUTURAS

O *Dashboard* concebido nesta tese pode ser considerado para futura incorporação institucional pela UFAM como ferramenta de apoio à gestão do portfólio tecnológico, no âmbito de suas estruturas administrativas de inovação. Sua atualização periódica (anual ou semestral) pode oferecer subsídios para observar a evolução dos 55 ativos identificados, monitorar novos depósitos de patentes e analisar possíveis mudanças na maturidade tecnológica das pesquisas à medida que avançam para estágios TRL 4-6.

Além de apoiar a gestão interna, o *Dashboard* pode atuar como ponto de convergência entre universidade, setor produtivo, governo e comunidades tradicionais. Ao disponibilizar dados estratégicos de modo transparente, ele contribui para reduzir assimetrias de informação e apoiar a formação de parcerias informadas por dados. A Cooperação Aberta, mediada por inteligência de dados, podendo contribuir para organizar a inovação como processo colaborativo e contínuo.

Os resultados desta tese também evidenciam um ponto estrutural relevante: indica um movimento de transição de práticas centradas em registros para abordagens analíticas baseadas em inteligência de dados. Enquanto métodos tradicionais de avaliação da inovação dependem de buscas manuais, indicadores restritos ou análises pontuais, as técnicas de NLP e ML indicaram capacidade de identificar padrões, maturidade tecnológica e oportunidades de proteção intelectual de maneira sistemática, reproduzível e escalável. Essa mudança de paradigma é coerente com as métricas de terceira geração o que pode ampliar a capacidade analítica dos NITs.

Nesse sentido, a tese oferece elementos que podem informar políticas públicas de CT&I na Amazônia: (i) a importância da proteção intelectual precoce, dada a baixa correlação entre TRL e patenteabilidade; (ii) importância de processos contínuos de qualificação das equipes dos NITs capacitadas em valoração, negociação e prospecção; e (iii) o potencial da IA para orientar a bioeconomia, reduzindo custos e aumentando a eficiência da bioprospecção.

A proposta alinha-se às metas do PDI 2016-2025 da UFAM, especialmente no eixo de empreendedorismo, tecnologia e inovação, e pode subsidiar programas voltados ao desenvolvimento regional.

A adoção institucional de modelos integrados de inteligência de dados, representa um caminho promissor para a modernização da política de inovação no país, especialmente no fortalecimento dos NITs e na articulação entre pós-graduação e transferência tecnológica.

Por fim, esta tese não se encerra com a defesa. O conceito de *Alive Thesis*, uma produção científica viva, em constante atualização e conectada às dinâmicas reais do território, indica que este trabalho pode ser compreendido como processo em movimento, articulado com agentes reais e relacionado às dinâmicas dos ecossistemas de pesquisa, inovação e políticas públicas da região.

REFERÊNCIAS

AHMAD, Anis Mardiana; KHURIZAN, Noor Saifurina Nana. A systematic review on data envelopment analysis application in higher education. *In: THE 15TH ASIAN CONGRESS ON BIOTECHNOLOGY*, 2024, Bali, Indonesia. **Anais [...]**. Bali, Indonesia: [s. d.], 2024. p. 070004. Disponível em: <https://pubs.aip.org/aip/acp/article-lookup/doi/10.1063/5.0192327>. Acesso em: 13 abr. 2026.

ALBA, Camila Fernanda; SUGUIMOTO, Hélio Hiroshi; MORIOKA, Luiz Rodrigo Ito. Prospecção tecnológica de patentes sobre compostos bioativos de microalgas / Technological prospecting of patents on microalgae bioactive compounds. **Brazilian Journal of Development**, [s. l.], v. 7, n. 8, p. 81223–81236, 2021.

ALLOUN, Wiem; CALVIO, Cinzia. Bio-Driven Sustainable Extraction and AI-Optimized Recovery of Functional Compounds from Plant Waste: A Comprehensive Review. **Fermentation**, [s. l.], v. 10, n. 3, p. 126, 2024.

ALMALKI, Hameeda A.; DURUGBO, Christopher M. Systematic review of institutional innovation literature: towards a multi-level management model. **Management Review Quarterly**, [s. l.], v. 73, n. 2, p. 731–785, 2023.

ALMEIDA, Robson Lopes De; MARICATO, João De Melo. Explorando conceitos e métricas de inovação no contexto das universidades. **Informação & Informação**, [s. l.], v. 26, n. 2, p. 646, 2021.

ALVES, Marco Antônio Baleeiro *et al.* Transferência de Tecnologia, Patentes e Inovação na Universidade Federal do Tocantins: um Estudo de Caso. **Cadernos de Prospecção**, [s. l.], v. 12, n. 5, p. 1257–1276, 2018.

AMARAL, Creusa Sayuri Tahara *et al.* Novos caminhos da biotecnologia: As inovações da indústria 4.0 na saúde humana. **Revista Brasileira Multidisciplinar**, [s. l.], v. 23, n. 3, p. 203–231, 2020.

AMBULI, T.V *et al.* AI-Driven Financial Management Optimizing Investment Portfolios through Machine Learning. *In: 2024 7TH INTERNATIONAL CONFERENCE ON CIRCUIT POWER AND COMPUTING TECHNOLOGIES (ICCPCT)*, 2024, Kollam, India. **2024 7th International Conference on Circuit Power and Computing Technologies (ICCPCT)**. Kollam, India: IEEE, 2024. p. 1822–1828. Disponível em: <https://ieeexplore.ieee.org/document/10672859/>. Acesso em: 21 jun. 2025.

AMRY, Dwitya K.; AHMAD, Ali J.; LU, Dawei. The new inclusive role of university technology transfer: Setting an agenda for further research. **International Journal of Innovation Studies**, [s. l.], v. 5, n. 1, p. 9–22, 2021.

ANDERSEN, Per; PETERSEN, Niels Christian. A Procedure for Ranking Efficient Units in Data Envelopment Analysis. **Management Science**, [s. l.], v. 39, n. 10, p. 1261–1264, 1993.

ANDERSON, Robert P. *et al.* **Ecological Niches and Geographic Distributions (MPB-49)**: [S. l.]: Princeton University Press, 2011. Disponível em:

<https://www.degruyterbrill.com/document/doi/10.1515/9781400840670/html>. Acesso em: 8 jul. 2026.

ARIMA JÚNIOR, Mauro Kiithi. Observações sobre as prioridades da Cooperação Amazônica: um esboço de políticas orientadas à missão no desenvolvimento da bioeconomia regional. **Revista Tempo do Mundo (RTM)**: n. 27, dez. 2021, [s. l.], v. 27, p. 165–189, 2022.

AVENI, Alessandro; CARVALHO, Sônia Marise Salles. Avaliação De Patentes E Inovação. Métodos E Problemas. **Cadernos De Prospecção**, [s. l.], v. 10, n. 4, p. 639, 2017.

AWASTHY, Richa *et al.* A framework to improve university–industry collaboration. **Journal of Industry-University Collaboration**, [s. l.], v. 2, n. 1, p. 49–62, 2020.

BANERJEE, Mousumi *et al.* Beyond kappa: A review of interrater agreement measures. **Canadian Journal of Statistics**, [s. l.], v. 27, n. 1, p. 3–23, 1999.

BANKER, R. D.; CHARNES, A.; COOPER, W. W. Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. **Management Science**, [s. l.], v. 30, n. 9, p. 1078–1092, 1984.

BARTZ, Cátia Raquel Felden *et al.* Colaboração e open innovation: a importância da governança colaborativa para a constituição de um ecossistema de inovação aberta em um Arranjo Produtivo Local (APL). **Interações (Campo Grande)**, [s. l.], p. 155–172, 2020.

BASHIR, Naheed. Knowledge identification for open innovation: a case study of biotechnology firms. **Business Process Management Journal**, [s. l.], v. 30, n. 4, p. 1131–1153, 2024.

BAUER, Ingrid *et al.* **Next Generation Metrics**. [S. l.]: Zenodo, 2020. Disponível em: <https://zenodo.org/record/3874801>. Acesso em: 28 mar. 2026.

BAXTER, Pamela; JACK, Susan. Qualitative Case Study Methodology: Study Design and Implementation for Novice Researchers. **The Qualitative Report**, [s. l.], 2015. Disponível em: <https://nsuworks.nova.edu/tqr/vol13/iss4/2/>. Acesso em: 12 maio 2025.

BEJARANO, José Bestier Padilla *et al.* Open innovation: A technology transfer alternative from universities. A systematic literature review. **Journal of Open Innovation: Technology, Market, and Complexity**, [s. l.], v. 9, n. 3, p. 1–12, 2023.

BOARDMAN, Anthony E *et al.* **Cost-benefit Analysis : Concepts and Practice**. [S. l.]: Cambridge University Press, 2018.

BOCHI, Fernanda; MOURA, Ana Maria Mielniczuk. Patentes em células-tronco: suas principais características e análise de citação. **Em Questão**, [s. l.], v. 26, n. 2, p. 32–57, 2020.

BRASIL. **Decreto nº 9.283 de 07 de fevereiro de 2018**. [S. l.]: D.O.U. DE 08/02/2018, p. 10, 2018. Disponível em:

<https://legislacao.presidencia.gov.br/atos/?tipo=DEC&numero=9283&ano=2018&ato=56elzZq5UeZpWTa46>.

BRASIL. **Lei 13.123 de 20 de maio de 2015 (Lei da Biodiversidade)**. [S. l.]: D.O.U de 21/05/2015, p. 1, 2015. Disponível em: <https://legislacao.presidencia.gov.br/atos/?tipo=LEI&numero=13123&ano=2015&ato=9a0ITU65UNVpWTc7b>.

BREIMAN, Leo. Random Forests. **Machine Learning**, Netherlands: Kluwer Academic Publishers, v. 45, n. 1, p. 5–32, 2001.

BRITT, Barry L. *et al.* Document classification techniques for automated technology readiness level analysis. **Journal of the American Society for Information Science and Technology**, [s. l.], v. 59, n. 4, p. 675–680, 2008.

CAMARA, Marcia Regina Gabardo *et al.* Análise da evolução do sistema nacional de inovação em biotecnologia no Brasil. **Gestão e Desenvolvimento em Revista**, [s. l.], v. 1, n. 1, p. 34–49, 2015.

CARAYANNIS, Elias G; BARTH, Thorsten D; CAMPBELL, David Fj. The Quintuple Helix innovation model: global warming as a challenge and driver for innovation. **Journal of Innovation and Entrepreneurship**, [s. l.], v. 1, n. 1, p. 2, 2012.

CARAYANNIS, Elias G.; PREISLER, Steffen. The quintuple innovation helix plus AI: A prospective retrospective approach and the use case of semiconductors. **Technological Forecasting and Social Change**, [s. l.], v. 216, p. 124122, 2025.

CARDOSO, Rayssa L. *et al.* The use of fisheries resources in the Brazilian patent system. **Anais da Academia Brasileira de Ciencias**, [s. l.], v. 94, n. 2, p. 1–20, 2022.

CARDOSO, Marcia Regina Gonçalves; OLIVEIRA, Guilherme Saramago De; GHELLI, Kelma Gomes Mendonça. Análise De Conteúdo: Uma Metodologia De Pesquisa Qualitativa. **Cadernos da Fucamp**, [s. l.], v. 20, n. 43, p. 98–111, 2021.

CAVALCANTE, Gabriel Martins. **Diretrizes para Plataformas Voltadas para a Transferência de tecnologias**. 2022. 1–98 f. PhD Thesis[s. l.], 2022. Disponível em: <https://tede.ufam.edu.br/handle/tede/10328>.

CHAI, T.; DRAXLER, R. R. Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. **Geoscientific Model Development**, [s. l.], v. 7, n. 3, p. 1247–1250, 2014.

CHAN, W. K. C. *et al.* Asset Optimization Using Integrated Modeling, Optimization, and Artificial Intelligence. [s. l.], 2019. Disponível em: <https://patents.google.com/patent/US20200387818A1/en>.

CHANG, Geba Jisung. Biodiversity estimation by environment drivers using machine/deep learning for ecological management. **Ecological Informatics**, [s. l.], v. 78, p. 102319, 2023.

CHANG, Ha-joon. **Chutando a escada: a estratégia do desenvolvimento em perspectiva histórica**. São Paulo: Unesp, 2004.

CHARNES, A.; COOPER, W. W.; RHODES, E. Measuring the efficiency of decision making units. **European Journal of Operational Research**, [s. l.], v. 2, n. 6, p. 429–444, 1978.

CHAVES, Maria do Perpétuo Socorro Rodrigues. A Gestão da Inovação na Universidade Federal do Amazonas. *In: BOAS PRÁTICAS DE GESTÃO EM NÚCLEOS DE INOVAÇÃO TECNOLÓGICA: EXPERIÊNCIAS INOVADORAS: VOLUME 1*. 1. ed. Jundiaí/SP: Edições Brasil, 2018. (Gestão da inovação em NIT). v. 1.

CHAVES, Ma. do P. Socorro Rodrigues. INOVAÇÃO E O APROVEITAMENTO DE FONTES LOCAIS DE CONHECIMENTO NA AMAZÔNIA: Desafios De Inclusão Social E Sustentabilidade. *In: A NOVA GERAÇÃO DE POLÍTICAS DE DESENVOLVIMENTO PRODUTIVO, SUSTENTABILIDADE SOCIAL E AMBIENTAL*. Brasília: [s. d.], 2012. p. 33–54.

CHIKKAMATH, Renuksamy *et al.* Patent Classification Using BERT-for-Patents on USPTO. *In: MLNLP 2022: 2022 5TH INTERNATIONAL CONFERENCE ON MACHINE LEARNING AND NATURAL LANGUAGE PROCESSING*, 2022, Sanya China. **Proceedings of the 2022 5th International Conference on Machine Learning and Natural Language Processing**. Sanya China: ACM, 2022. p. 20–28. Disponível em: <https://dl.acm.org/doi/10.1145/3578741.3578746>. Acesso em: 18 nov. 2025.

CHOUBEY, Sachin; KARMAKAR, G. P. Artificial Intelligence Techniques and Their Application in Oil and Gas Industry. **Artificial Intelligence Review**, [s. l.], v. 54, n. 54, 2020.

COHEN, Jacob. A Coefficient of Agreement for Nominal Scales. **Educational and Psychological Measurement**, [s. l.], v. 20, n. 1, p. 37–46, 1960.

COHEN, Maria De Los Angeles Martinez; PAIXÃO, Adriano Nascimento; OLIVEIRA, Nilton Marques. Eficiência Nas Universidades Federais Brasileira: Uma Aplicação Da Análise Envoltória De Dados. **Informe GEPEC**, [s. l.], v. 22, n. 1, p. 133–149, 2018.

COLL-SERRANO, V.; BOLÓS, V. J.; SUÁREZ, Benítez R. **deaR: Data Envelopment Analysis in R**. versão 1.3.2. [S. l.]: [s. d.], 2022 [2025]. Disponível em: <https://github.com/ElsevierSoftwareX/SOFTX-D-25-00164>. Acesso em: 17 maio 2026.

COMPAGNUCCI, Lorenzo; SPIGARELLI, Francesca. The Third Mission of the university: A systematic literature review on potentials and constraints. **Technological Forecasting and Social Change**, [s. l.], v. 161, p. 120284, 2020.

CONGALTON, Russell G. A review of assessing the accuracy of classifications of remotely sensed data. **Remote Sensing of Environment**, [s. l.], v. 37, n. 1, p. 35–46, 1991.

COOK, Wade D.; ZHU, Joe. **Data Envelopment Analysis: Modeling Operational Processes and Measuring Productivity**. s.l.: [s. d.], 2008 [2008].

COOPER, William Wager; SEIFORD, Lawrence M.; TONE, Kaoru. **Data envelopment analysis: a comprehensive text with models, applications, references and DEA-solver software**. 2nd eded. New York: Springer, 2007. (Springer e-books).

CORRÊA, Alessandro de Castro *et al.* Análise Envolvória De Dados Com Apoio Do Software R: Uma Aplicação Nos Institutos Federais. **Revista de Engenharia e Tecnologia**, [s. l.], v. 11, n. 2, p. 1–13, 2019.

COSTA, Carlos A. Bana e; MELLO, João Carlos C. B. Soares de; MEZA, Lidia Angulo. A new approach to the bi-dimensional representation of the DEA efficient frontier with multiple inputs and outputs. **European Journal of Operational Research**, [s. l.], v. 255, n. 1, p. 175–186, 2016.

COUTINHO, Tamires L.; ANDRADE, Leandro; SANTANA, Cleber Jorge L. De. Soluções e desafios: métodos ágeis na pesquisa colaborativa entre universidade e indústria. *In*: ANAIS ESTENDIDOS DO SIMPÓSIO BRASILEIRO DE SISTEMAS COLABORATIVOS, 2024, Brasil. **Anais Estendidos do XIX Simpósio Brasileiro de Sistemas Colaborativos (SBSC Estendido 2024)**. Brasil: Sociedade Brasileira de Computação - SBC, 2024. p. 247–251. Disponível em: https://sol.sbc.org.br/index.php/sbsc_estendido/article/view/28166. Acesso em: 26 jun. 2025.

CUTLER, D. Richard *et al.* Random Forests for Classification in Ecology. **Ecology**, [s. l.], v. 88, n. 11, p. 2783–2792, 2007.

D'AMATO, D.; KORHONEN, J. Integrating the green economy, circular economy and bioeconomy in a strategic sustainability framework. **Ecological Economics**, [s. l.], p. 12, 2021.

DANACH, Kassem; DIRANI, Ali El; RKEIN, Hassan. Revolutionizing Supply Chain Management With AI: A Path to Efficiency and Sustainability. **IEEE Access**, [s. l.], v. 12, p. 188245–188255, 2024.

DASGUPTA, Partha. **The economics of biodiversity: the Dasgupta review: full report**. Updated: 18 February 2021ed. London: HM Treasury, 2021.

DE LAS HERAS-ROSAS, Carlos; HERRERA, Juan. Research Trends in Open Innovation and the Role of the University. **Journal of Open Innovation: Technology, Market, and Complexity**, [s. l.], v. 7, n. 1, p. 29, 2021.

DE MEDEIROS, Cassandra Carneiro; PELAEZ, Victor. Papel Do Inpi No Processo De Institucionalização Da Propriedade Industrial No Brasil. **Gestão e Sociedade**, [s. l.], v. 15, n. 43, p. 4411–4435, 2021.

DE MENEZES, Cristiana Souza *et al.* Desafios e oportunidades para a transferência de tecnologia nas instituições científicas, tecnológicas e de inovação públicas brasileiras: uma análise a partir dos núcleos de inovação tecnológica. **Revista JRG de Estudos Acadêmicos**, [s. l.], v. 8, n. 19, p. e082805, 2025.

DU, Yamin; SEO, Wonchul. A Comparative Study on the Efficiency of R&D Activities of Universities in China by Region Using DEA–Malmquist. **Sustainability**, [s. l.], v. 14, n. 16, p. 10433, 2022.

DYSON, R.G. *et al.* Pitfalls and protocols in DEA. **European Journal of Operational Research**, [s. l.], v. 132, n. 2, p. 245–259, 2001.

EL GIBARI, Samira; GÓMEZ, Trinidad; RUIZ, Francisco. Combining reference point based composite indicators with data envelopment analysis: application to the assessment of universities. **Scientometrics**, [s. l.], v. 127, n. 8, p. 4363–4395, 2022.

ELITH, Jane; LEATHWICK, John R. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. **Annual Review of Ecology, Evolution, and Systematics**, [s. l.], v. 40, n. 1, p. 677–697, 2009.

EMROUZNEJAD, Ali; YANG, Guo-liang. A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016. **Socio-Economic Planning Sciences**, [s. l.], v. 61, p. 4–8, 2018.

EUROPEAN COMMISSION. High-level summary of the AI Act | EU Artificial Intelligence Act. *In*: 2024. Disponível em: <https://artificialintelligenceact.eu/high-level-summary/>. Acesso em: 19 maio 2026.

FAHEEM, Muhammad Ashraf; ASLAM, Muhammad; KAKOLU, Sridevi. Artificial Intelligence in Investment Portfolio Optimization: A Comparative Study of Machine Learning Algorithms. **International Journal of Science and Research Archive**, [s. l.], v. 6, n. 1, p. 335–342, 2022.

FALVO, David; WEIDENER, Lukas; KARLSSON, Martin. **Developing ERAF-AI: An Early-Stage Biotechnology Research Assessment Framework Optimized For Artificial Intelligence Integration**. [S. l.]: Bioinformatics, 2025. Disponível em: <http://biorxiv.org/lookup/doi/10.1101/2025.01.08.631843>. Acesso em: 4 jun. 2026.

FÁVERI, José Ernesto De. Fundamentos para pensar a inovação sustentável. **Interação - Revista de Ensino, Pesquisa e Extensão**, [s. l.], v. 22, n. 2, p. 49–63, 2020.

FERNANDES, Daniela Rangel Affonso; GADELHA, Carlos Augusto Grabois; MALDONADO, Jose Manuel Santos de Varge. Vulnerabilidades das indústrias nacionais de medicamentos e produtos biotecnológicos no contexto da pandemia de COVID-19. **Cadernos de Saúde Pública**, [s. l.], v. 37, n. 4, p. 1–14, 2021.

FERREIRA, Douglas Gomes. Análise De Dados Em Linguagem R. **Enap.gov.br**, [s. l.], 2020. Disponível em: <http://repositorio.enap.gov.br/handle/1/7872>. Acesso em: 1 jan. 2025.

FERREIRA, Maria Paula da Rosa; TERRA, Rosane Beatris Mariano da Rocha Barcellos; GREGORI, Isabel Christine Silva De. Patentes e sementes transgênicas: o viés econômico-mercado. **Revista de Direito Econômico e Socioambiental**, [s. l.], v. 11, n. 3, p. 81–109, 2020.

FILGUEIRAS, Fernando; LUI, Lizandro. Designing data governance in Brazil: an institutional analysis. **Policy Design and Practice**, [s. l.], v. 6, n. 1, p. 41–56, 2023.

FIRSOVA, Anna; CHERNYSHOVA, Galina. Efficiency Analysis of Regional Innovation Development Based on DEA Malmquist Index. **Information**, [s. l.], v. 11, n. 6, p. 294, 2020.

FLORENCIO, Márcio Nannini da Silva *et al.* Análise da produção e colaboração da biotecnologia no Brasil. **Research, Society and Development**, [s. l.], v. 9, n. 7, p. 1–27, 2020.

FONTES, André C. R. Transferência de biotecnologia. **Revista Interdisciplinar do Direito - Faculdade de Direito de Valença**, [s. l.], v. 18, n. 2, p. 12, 2020.

FOODY, Giles M. Status of land cover classification accuracy assessment. **Remote Sensing of Environment**, [s. l.], v. 80, n. 1, p. 185–201, 2002.

FRANKLIN, Janet; MILLER, Jennifer A. **Mapping species distributions: spatial inference and prediction**. Cambridge ; New York: Cambridge University Press, 2009. (Ecology, biodiversity and conservation).

GIANNINI, Tereza C. *et al.* Desafios Atuais Da Modelagem Preditiva De Distribuição De Espécies. **Rodriguésia**, [s. l.], v. 63, n. 3, p. 733–749, 2012.

GROLEMUND, Garrett. **RStudio**. versão 3.0.3. [S. l.]: [s. d.], 2018. Disponível em: <https://rstudio-education.github.io/hopr/>. Acesso em: 20 mar. 2025.

GUALANDI FILHO, Paulo Elias *et al.* Avaliação de eficiência de universidades federais brasileiras: uma abordagem pela Análise Envoltória de Dados. **Avaliação: Revista da Avaliação da Educação Superior (Campinas)**, [s. l.], v. 28, p. e023018, 2023.

GUERRA, Avaetê de Lunetta e Rodrigues. Metodologia Da Pesquisa Científica E Acadêmica. **REVISTA OWL**, [s. l.], v. 1, n. 2, p. 149–159, 2023.

GUERRERO, Maribel; SIEGEL, Donald S. Prosocial Technology Transfer and Academic Entrepreneurship: Lessons Learned and New Directions. **Academy of Management Perspectives**, [s. l.], p. 2–18, 2024.

GUERRERO, Maribel; URBANO, David. Effectiveness of technology transfer policies and legislation in fostering entrepreneurial innovations across continents: an overview. **Journal of Technology Transfer**, [s. l.], v. 44, n. 5, p. 1347–1366, 2019.

GUPTA, Supriya; SHARMA, Abhishek; ABUBAKAR, Aria. Artificial Intelligence–Driven Asset Optimizer. *In*: SPE ANNUAL TECHNICAL CONFERENCE AND EXHIBITION, 2018, Dallas, Texas, USA. **SPE Annual Technical Conference and Exhibition**. Dallas, Texas, USA: SPE, 2018. p. D012S045R001. Disponível em: <https://onepetro.org/SPEATCE/proceedings/18ATCE/18ATCE/D012S045R001/213656>. Acesso em: 21 jun. 2025.

HAN, Sunwoo; WILLIAMSON, Brian D.; FONG, Youyi. Improving random forest predictions in small datasets from two-phase sampling designs. **BMC Medical Informatics and Decision Making**, [s. l.], v. 21, n. 1, p. 322, 2021.

HANNOUF, Marwa B. *et al.* Social life cycle assessment (S-LCA) of technology systems at different stages of development. **The International Journal of Life Cycle Assessment**, [s. l.], v. 30, n. 6, p. 1099–1114, 2025.

HASAN, Sobuj; SHAFIQ, Sadman Bin; KHATUN, Lubna. EXPLORING THE POTENTIAL OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING IN

MINERAL EXPLORATION: A REVIEW ARTICLE. **International Research Journal of Modernization in Engineering Technology and Science**, [s. l.], v. 5, n. 10, 2023.

HASSANI, Hossein *et al.* Text Mining in Big Data Analytics. **Big Data and Cognitive Computing**, [s. l.], v. 4, n. 1, p. 1, 2020.

HIDALGO, Enric Senabre. Management of a Multidisciplinary Research Project: A Case Study on Adopting Agile Methods. **Journal of Research Practice**, [s. l.], v. 14, n. 1, p. 1–15, 2018.

HOLGERSSON, Marcus; AABOEN, Lise. A literature review of intellectual property management in technology transfer offices: From appropriation to utilization. **Technology in Society**, [s. l.], v. 59, n. 2018, p. 1–10, 2019.

HOLZINGER, Andreas *et al.* AI for life: Trends in artificial intelligence for biotechnology. **New Biotechnology**, [s. l.], v. 74, p. 16–24, 2023.

INEP. **Censo da Educação Superior**. [S. l.], 2020. Disponível em: <https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/censo-da-educacao-superior/censo-da-educacao-superior>. Acesso em: 21 mar. 2025.

INPI. **IBID: O mapa da inovação do Brasil em suas mãos**. Rio de Janeiro: INPI, 2025. Disponível em: https://www.gov.br/inpi/pt-br/inpi-data/indice-brasil-de-inovacao-e-desenvolvimento-ibid/ibid_2025_pt-br-final.pdf. Acesso em: 6 jul. 2026.

IPIRANGA, Ana Sílvia Rocha; CHAYM, Carlos Dias; SOUSA, Felipe Gerhard Paula. Relatos Sobre O Organizar Do Sócio-Passado Em Uma Patente Brasileira De Biotecnologia. **Revista Eletrônica de Ciência Administrativa**, [s. l.], v. 15, n. 2, p. 133–147, 2016.

ISO. **Sistema de Gestão de Inteligência Artificial (SGAI)**. Geneva: International Organization for Standardization, , 2023.

JAIN, Bhavesh Mahender; KUMAR, Deepak. Predicting Tech Readiness through Bibliometric Analysis using Unsupervised Machine Learning. [s. l.], 2025.

JIA, Lihua *et al.* Development of interactive biological web applications with R/Shiny. **Briefings in Bioinformatics**, [s. l.], v. 23, n. 1, p. bbab415, 2022.

JOEL, Olorunyomi Stephen *et al.* Leveraging Artificial Intelligence for Enhanced Supply Chain Optimization: A Comprehensive Review of Current Practices and Future Potentials. **International Journal of Management & Entrepreneurship Research**, [s. l.], v. 6, n. 3, p. 707–721, 2024.

JOHNSON, Ben. AI biotechs launch bioprospecting expeditions with Indigenous groups, agree to share benefits. **Nature Biotechnology**, [s. l.], v. 43, n. 2, p. 154–156, 2025.

KARAMITROU, Alexandra *et al.* Towards the Use of Artificial Intelligence Deep Learning Networks for Detection of Archaeological Sites. **IOP Publishing**, [s. l.], v. 10, n. 4, p. 044001–044001, 2022.

KAUR, Gurpreet; MALIK, Kamal. A Comprehensive Overview of Sentiment Analysis and Fake Review Detection. *In: LECTURE NOTES IN NETWORKS AND SYSTEMS*, 2020. **Anais [...]**. [S. l.]: Springer International Publishing, 2020. p. 293–304.

KAWABATA, Marcelo Koji; CAMARGO JUNIOR, Alceu Salles. Innovation and institutions' quality: a comparative study between countries. **International Journal of Innovation Science**, [s. l.], v. 12, n. 2, p. 169–185, 2020.

KHEZRIMOTLAGH, Dariush *et al.* Data Envelopment Analysis and Big Data. **European Journal of Operational Research**, [s. l.], v. 274, n. 3, p. 1047–1054, 2019.

KIM, Jootae; ICK, Jin. Measuring Institutional Efficiency and Innovation Efficiency: Case of Korea and ASEAN. 2019, Rennes, France. **XVIII International Business & Economy Conference (IBEC)**. Rennes, France: [s. d.], 2019.

KISSIMOTO, Kumiko Oshio; MATTOS, Cláudia Aparecida; LAURINDO, Fernando José. The role of the organizational and operational dimensions in the open collaboration performance: a strategic alignment perspective. **RAUSP Management Journal**, [s. l.], v. 58, n. 1, p. 22–38, 2023.

KÜÇÜKDEMIRCI, Melda; SARRIS, Apostolos. GPR Data Processing and Interpretation Based on Artificial Intelligence Approaches: Future Perspectives for Archaeological Prospection. **Remote Sensing**, [s. l.], v. 14, n. 14, p. 3377, 2022.

KUHN, Max; JOHNSON, Kjell. **Applied predictive modeling**. Corrected at 5th printing. New York: Springer, 2016.

LANDIS, J. Richard; KOCH, Gary G. The Measurement of Observer Agreement for Categorical Data. **Biometrics**, [s. l.], v. 33, n. 1, p. 159, 1977.

LANGER, Sarah. The relation of standards and power in management and organization research: Core notions and alternative avenues. **International Journal of Management Reviews**, [s. l.], v. 25, n. 4, p. 647–665, 2023.

LATIPAT. **Latipat-Espacenet**. [S. l.], 2017. Disponível em: https://lp.espacenet.com/?locale=pt_LP. Acesso em: 2 fev. 2025.

LAW INSIDER. **Public benefit corporation Definition: 146 Samples**. [S. l.], 2025. Disponível em: <https://www.lawinsider.com/dictionary/public-benefit-corporation>. Acesso em: 13 abr. 2026.

LEÃO, Delfim *et al.* **Comunidades de Ciência Aberta na UC 2024: Livro de resumos**. [S. l.]: [s. d.], 2024. Disponível em: <https://hdl.handle.net/10316/115518>.

LEE, Heejae *et al.* Optimal Training Sample Sizes for U-Net-Based Tree Species Classification with Sentinel-2 Imagery. **Forests**, [s. l.], v. 16, n. 11, p. 1718, 2025.

LESSA, Edison Souza; DE SOUZA, André Luís Rocha; SILVA, Maria Valesca Damásio De Carvalho. DESAFIOS NA CONTABILIZAÇÃO DE ATIVOS DE PROPRIEDADE INTELECTUAL NO ÂMBITO DAS ICTs PÚBLICAS FEDERAIS BRASILEIRA. **ARACÊ**, [s. l.], v. 6, n. 4, p. 14400–14426, 2024.

LETTI, Ariel Gustavo; BITTENCOURT, Mauricio Vaz Lobo; VILA, Luis E. Análise Comparativa Da Eficiência Das Universidades Federais Entre as Regiões Brasileiras (2010 a 2016). [s. l.], v. 16, n. 1, p. 382–396, 2020.

LI, Shi; CHEN, Jianping; LIU, Chang. Overview on the Development of Intelligent Methods for Mineral Resource Prediction under the Background of Geological Big Data. **Minerals**, [s. l.], v. 12, n. 5, p. 616, 2022.

LIANG, Qingqing; YIN, Fang. **Research Efficiency Analysis of Universities in 31 Chinese Provinces Using Three-Stage DEA Model and Dynamic Malmquist Index-Based on Provincial Panel Data from 2012 to 2022**. [S. l.]: Social Sciences, 2024. Disponível em: <https://www.preprints.org/manuscript/202406.0596/v1>. Acesso em: 20 mar. 2025.

LIBOREIRO, Karla Rocha. **Interação universidade-empresa em biotecnologia: estudos de caso em laboratórios de pesquisa universitários estadunidenses e brasileiros**. 2020. 1–223 f. [s. l.], 2020. Disponível em: <http://hdl.handle.net/1843/36104>.

LIMA, Livia Maria Queiroz *et al.* Reflexões Sobre o Uso de Indicadores de Propriedade Intelectual na Análise de Eficiência dos NITS. **Cadernos de Prospecção**, [s. l.], v. 16, n. 1, p. 144–160, 2023.

LOPES, Sânya Léa Alves Rocha. **Avaliação da gestão de transferência de tecnologia nas instituições científicas, tecnológicas e de inovação do Brasil**. 2019. 129 f. Dissertação - UnB, Brasília - DF, 2019. Disponível em: <http://repositorio.unb.br/handle/10482/38046>.

LOUREIRO, Violeta. **Caminhos e descaminhos da Amazônia em busca de desenvolvimento**. Manaus, AM: Editora Valer, 2023. v. 1 e 2

LOVETH, Roscoe Goble. **Emerging Technologies in AI-Driven Biotech Innovations**. [S. l.], 2025. Disponível em: https://www.researchgate.net/publication/389465134_Emerging_Technologies_in_AI-Driven_Biotech_Innovations#fullTextFileContent. .

LU, Tan; YAN BAI, Wang. The Study of the DEA Method Model for University Teaching Quality Assessment and Benefit Analysis. **Physics Procedia**, [s. l.], v. 25, p. 1187–1191, 2012.

MARCHETTI, Ivanir *et al.* Interação Universidade E Empresa: Barreiras E Desafios Na Transferência De Tecnologia. **Revista Estudo & Debate**, [s. l.], v. 31, n. 1, 2024. Disponível em: <https://univates.br/revistas/index.php/estudoedebate/article/view/3503>. Acesso em: 12 dez. 2025.

MARINHO, Enrico Picoli; OLIVEIRA, Deyla Paula De; ARAÚJO, Gustavo Cunha De. Propriedade Intelectual nos cursos de Biotecnologia das Universidades Federais do Brasil. **Ensino em Re-Vista**, [s. l.], v. 30, p. e019, 2023.

MARKUS, Monica Th.; GROENEN, Patrick J. F. B. Efron and R. J. Tibshirani (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall, xvi + 436 pp. ISBN 0-412-0423t-2, \$50.00. **Psychometrika**, [s. l.], v. 63, n. 1, p. 97–101, 1998.

MARQUES, Humberto Rodrigues *et al.* Inovação Aberta entre Universidade-Empresa: a Percepção de Professores Universitários. **Sociedade, Contabilidade e Gestão**, [s. l.], v. 16, n. 1, p. 83–104, 2021.

MCTI. Política de Integridade na Atividade Científica do CNPq. **Portaria n. 2664 de 6 de março de 2026**, n. 2664, 2664, 6 mar. 2026. Disponível em: http://memoria2.cnpq.br/web/guest/view/-/journal_content/56_INSTANCE_0oED/10157/23142775.

MCTI FORTEC. **Guia de orientação: contratos de transferência de tecnologia nos termos do marco legal da ciência, tecnologia e inovação**. Brasília, DF: Ministério da Ciência, Tecnologia e Inovações, 2025. Disponível em: https://fortec.org.br/wp-content/uploads/2023/07/Guias-de-Orientac%CC%A7o%CC%83es-sobre-Instrumentos-do-Marco-Legal-de-CTI-CONTRATOS-DE-TRANSFERENCIA-DE-TECNOLOGIA_com-Prefacio_A4_Fortec_rev-MCTI.pdf.

MEDIANSYAH, Iski; SEPTIAN, Firza; ZIKRY, Arief. Penerapan Whale Optimization Algorithm dalam Pengoptimalan Portofolio Investasi Menggunakan Model Prediktif Artificial Intelligence. **Jurnal Software Engineering and Computational Intelligence**, [s. l.], v. 2, n. 01, p. 50–58, 2024.

MELO, Hugo Marciano D E. **Aplicação Da Análise Envolvória De Dados (DEA) Para Avaliação De Desempenho Das Unidades Acadêmicas De Uma Universidade Federal**. 2019. [s. l.], 2019. Disponível em: <http://repositorio.bc.ufg.br/tede/handle/tede/9670>.

MENEZES, Cristiana Souza De *et al.* Valoração das patentes nas icts públicas brasileiras: acesso à informação como vetor de eficiência na política nacional de inovação. **Revista Políticas Públicas & Cidades**, [s. l.], v. 15, n. 1, p. e3028, 2026.

MIRANDA, Nathalia Viana De *et al.* Análise Comparativa Das Parcerias Universidade-Indústria : Um Estudo De Análise Comparativa Das Parcerias Universidade- Indústria : Um Estudo De Caso Entre A Ufam E A Ufmg . **IOSR Journal of Business and Management**, [s. l.], v. 26, n. 9, p. 59–66, 2024.

MISHRA, Manoranjan *et al.* Spatial analysis and machine learning prediction of forest fire susceptibility: a comprehensive approach for effective management and mitigation. **Science of The Total Environment**, [s. l.], v. 926, p. 171713, 2024.

MISHRA, Sanjay; TIWARI, Amit M. Current Trends of Artificial Intelligence in Biotechnology. **Open Access Journal of Data Science and Artificial Intelligence**, [s. l.], v. 2, n. 1, 2024. Disponível em: <https://medwinpublishers.com/OAJDA/current-trends-of-artificial-intelligence-in-biotechnology.pdf>. Acesso em: 6 nov. 2025.

MOJJADA, Harihararao. Protocol for Metadata Harvesting: The Role of OAI-PMH in Digital Resource Integration. **International Journal of Research and Innovation in Applied Science**, [s. l.], v. X, n. VII, p. 724–736, 2025.

MONIGHA, Adogioye. AI-Driven Cost Optimization in Oil and Gas Projects. **International Journal of Petroleum and Gas Engineering Research**, [s. l.], v. 7, n. 2, p. 17–32, 2024.

MONTEIRO, Arainy Fernandes dos Santos. **Parcerias público-privadas como ferramentas de gestão no sector da saúde**. 2022. 14 f. Dissertação - INSTITUTO POLITÉCNICO DE BRAGANÇA, Bragança, 2022. Disponível em: <http://hdl.handle.net/10198/28132>. Acesso em: 21 nov. 2025.

MORALES SÁNCHEZ, Mario Alberto *et al.* Tendencias tecnológicas en el sector biotecnológico: análisis de patentes en México y Estados Unidos. **Economía Teoría y Práctica**, [s. l.], n. 51, 2019. Disponível em: <http://economiatyp.uam.mx/index.php/ETYP/article/view/403>. Acesso em: 18 mar. 2025.

MOREIRA, Aldemir Freire *et al.* Indicadores de inovação do Brasil e decisão de investimentos em políticas públicas. **Cadernos de Ciência & Tecnologia**, Brasília, DF, v. 41, p. 27443, 2024.

NAJAFI, Seyed *et al.* Performance evaluation accounting with inputs non-discretionary factors in an integrated BSC-DEA methodology. **Big Data and Computing Visions**, [s. l.], v. 3, n. 3, 2023. Disponível em: <https://doi.org/10.22105/bdcv.2023.190170>. Acesso em: 21 nov. 2025.

NAZARKO, Joanicjusz; ŠAPARAUSKAS, Jonas. APPLICATION OF DEA METHOD IN EFFICIENCY EVALUATION OF PUBLIC HIGHER EDUCATION INSTITUTIONS. **Technological and Economic Development of Economy**, [s. l.], v. 20, n. 1, p. 25–44, 2014.

OCDE. **OECD Framework for the Classification of AI systems**: OECD Digital Economy Papers. [S. l.]: [s. d.], 2022. OECD Digital Economy Papers. Disponível em: https://www.oecd.org/en/publications/oecd-framework-for-the-classification-of-ai-systems_cb6d9eca-en.html. Acesso em: 19 maio 2026.

OCDE. **Oslo Manual 2018: Guidelines for Collecting, Reporting and Using Data on Innovation, 4th Edition**. [S. l.]: OECD Publishing, 2018. (The Measurement of Scientific, Technological and Innovation Activities). Disponível em: https://www.oecd.org/en/publications/oslo-manual-2018_9789264304604-en.html. Acesso em: 17 abr. 2026.

OCDE (org.). **The bioeconomy to 2030: designing a policy agenda**. Paris: Organization for Economic Co-operation and Development, 2009.

O'DWYER, Michele; FILIERI, Raffaele; O'MALLEY, Lisa. Establishing successful university–industry collaborations: barriers and enablers deconstructed. **The Journal of Technology Transfer**, [s. l.], v. 48, n. 3, p. 900–931, 2023.

OKIMOTO, Carolina Alves. **Transferência De Tecnologia E Gestão Da Propriedade Intelectual: Um Estudo De Caso Na Universidade De Brasília**. 2023. 206 f. - Universidade de Brasília, [s. l.], 2023. Disponível em: https://ppgi.unb.br/images/documentos/Dissertacoes/Carolina_Alves_Okimoto.pdf.

OMROW, Delon *et al.* Compliance and enforcement in a brave new (green) world: best practices and technologies for green governance. **FACETS**, [s. l.], v. 9, p. 1–8, 2024.

PAIVA, Adriana Borges de; OLIVEIRA, Guilherme Saramago de; HILLESHEIM, Mara Cristina Piolla. Análise De Conteúdo: Uma Técnica De Pesquisa Qualitativa. **Análise De Conteúdo: Uma Técnica De Pesquisa Qualitativa**, [s. l.], v. 2, n. 1, p. 16–33, 2021.

PAIVA, Teresa; RIBEIRO, Maximiano; COUTINHO, Paula. R&D Collaboration, Competitiveness Development, and Open Innovation in R&D. **Journal of Open Innovation: Technology, Market, and Complexity**, [s. l.], v. 6, n. 4, p. 116, 2020.

PALAVECINO, Louise Azulay. **Patentes na área farmacêutica e biotecnologia : Atividade Inventiva e Prospecção Tecnológica de Anticorpos Monoclonais (mAbs) e Construção de Banco de Dados com Combinação de Fármacos**. 2021. PhD Thesis - Universidade Federal Fluminense, [s. l.], 2021. Disponível em: <http://app.uff.br/riuff/handle/1/29025>.

PALETTA, Francisco Carlos; CHUBA, Thiago Negrão. Aspectos Gerais Da Transferência De Tecnologia Em Biotecnologia No Brasil: Uma Abordagem Da Ciência Da Informação. In: COSTA, Edwaldo; HARA, Suélen Keiko. **A produção do conhecimento nas ciências da comunicação 2**. 1. ed. [S. l.]: Atena Editora, 2022. p. 33–56. Disponível em: <https://www.atenaeditora.com.br/post-ebook/4889>. Acesso em: 24 mar. 2025.

PARANHOS, Julia *et al.* O financiamento à inovação para a indústria farmacêutica brasileira: estudo de caso dos programas da FINEP e BNDES no atendimento às prioridades de saúde. **Revista Econômica**, [s. l.], v. 23, n. 1, p. 73–100, 2021.

PATERNOSTRO, Andre Góes; QUINTELLA, Cristina M.; LEITE, Handerson Jorge Dourado. Pesquisa Exploratória Comparativa entre Artigos e Patentes Sobre Maturidade (Prontidão) Tecnológica. **Cadernos de Prospecção**, [s. l.], v. 13, n. 4, p. 1088, 2020.

PATIAS, Ivete Aparecida; BAGGIO, Daniel Knebel. Núcleo De Inovação Tecnológica E Desenvolvimento Regional: Estudo Em Uma Instituição Federal De Educação, Ciência E Tecnologia, a Partir Da Atuação Da Hélice Quádrupla. **Informe GEPEC**, [s. l.], v. 26, n. 3, p. 342–358, 2022.

PEREIRA, Denise Pinho. **Análise Da Eficiência Das Universidades Federais Brasileiras: Uma Aplicação Da Análise Envoltória De**. 2020a. 1–94 f. - Universidade Federal do Tocantins, PALMAS-TO, 2020.

PEREIRA, Denise Pinho. **Análise Da Eficiência Das Universidades Federais Brasileiras : Uma Aplicação Da Análise Envoltória De Dados**. 2020b. [s. l.], 2020. Disponível em: <http://hdl.handle.net/11612/1943>.

PIMENTA, Fabricia Pires. A patente como fonte de informação (des)necessária para a Biotecnologia em Saúde. **Transinformacao**, [s. l.], v. 29, n. 3, p. 323–332, 2017.

PINHEIRO, Roger Rosado *et al.* Avaliação De Desempenho Das Unidades Acadêmicas De Uma Universidade Federal Brasileira Por Meio De Indicadores De

Eficiência. **Desafio Online**, [s. l.], v. 11, n. 2, 2022. Disponível em: <https://desafioonline.ufms.br/index.php/deson/article/view/15422>. Acesso em: 20 mar. 2025.

PINTO, Christofer Alex Souza; DE MATOS, Murilo Montanari. O Impacto da Biotecnologia sobre as Novas Entidades Químicas. 2021. **V Encontro Nacional de Economia Industrial e Inovação**. [S. l.]: [s. d.], 2021. p. 1–16. Disponível em: <https://pdf.blucher.com.br/engineeringproceedings/v-enei/661.pdf>.

POMEROY, Brett; GRILC, Miha; LIKOZAR, Blaž. Artificial neural networks for bio-based chemical production or biorefining: A review. **Renewable and Sustainable Energy Reviews**, [s. l.], v. 153, p. 111748, 2022.

POSIT. About. *In*: POSIT. 2022. Disponível em: <https://posit.co/about/>. Acesso em: 13 abr. 2026.

POSIT PBC. RStudio is now Posit! *In*: POSIT. 2 nov. 2022. Disponível em: <https://posit.co/blog/rstudio-is-now-posit/>. Acesso em: 30 mar. 2026.

PRAJAPATI, Parixit *et al.* Deep Learning Approaches for Predicting Bioactivity of Natural Compounds. **The Natural Products Journal**, [s. l.], v. 15, 2025. Disponível em: <https://www.eurekaselect.com/238324/article>. Acesso em: 10 jun. 2025.

PUBMED. **PubMed**. [S. l.], 2019. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/>. Acesso em: 21 mar. 2025.

QUINTELLA, Cristina M. A REVISTA CADERNOS DE PROSPECÇÃO E OS NÍVEIS DE MATURIDADE DE TECNOLOGIAS (TRL). **Cadernos de Prospecção**, [s. l.], v. 10, n. 1, p. 1, 2017.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria: R Foundation for Statistical Computing, 2026. Disponível em: <https://cran.r-project.org/doc/manuals/r-release/fullrefman.pdf>. Acesso em: 19 mar. 2026.

RABELO, Kassia Cristina de Caldas *et al.* Patenteamento de biotecnologias no Brasil: proteção patentária no campo biotecnológico de angiospermas nativas brasileiras. **Cadernos de Ciência & Tecnologia**, [s. l.], v. 40, p. 1–19, 2023.

RAJ, Wahid Zaman. THE ROLE OF DATA SCIENCE IN OPTIMIZING PROJECT EFFICIENCY AND INNOVATION IN U.S. ENTERPRISES. **International Journal of Business and Economics Insights**, [s. l.], v. 05, n. 03, p. 586–600, 2025.

RANE, Nitin; CHOUDHARY, Saurabh; RANE, Jayesh. Artificial intelligence and machine learning for resilient and sustainable logistics and supply chain management. **SSRN Electronic Journal**, [s. l.], 2024. Disponível em: <https://www.ssrn.com/abstract=4847087>. Acesso em: 21 jun. 2025.

RIBEIRO, Gislane Oliveira *et al.* Innovations and developments in single cell protein: Bibliometric review and patents analysis. **Frontiers in Microbiology**, [s. l.], p. 1–16, 2023.

RODRIGUES, Ana Luísa *et al.* A IMPORTÂNCIA DA RELAÇÃO ENTRE AS INSTITUIÇÕES DE ENSINO SUPERIOR E A SOCIEDADE – COLABORAÇÃO, INOVAÇÃO E TRANSFERÊNCIA DE CONHECIMENTO / THE RELATIONSHIP BETWEEN HIGHER EDUCATION INSTITUTIONS AND SOCIETY – COLLABORATION, INNOVATION AND TRANSFER OF KNOWLEDGE. **Brazilian Journal of Development**, [s. l.], v. 6, n. 12, p. 96491–96498, 2020.

ROSA, Luana Maria Tavares; SABONARO, Débora Zumkeller. A Importância Da Diversidade Biológica Na Política E Patentes Atrelada Aos Processos Biotecnológicos. **Revista da Universidade Vale do Rio Verde**, [s. l.], v. 17, n. 2, p. 1–10, 2019.

RUF: RANKING UNIVERSITÁRIO FOLHA – MELHORES UNIVERSIDADES, INSTITUIÇÕES E CURSOS 2025 | FOLHA | FOLHA. [S. l.], 2025. Disponível em: <https://ruf.folha.uol.com.br/2025>. Acesso em: 20 mar. 2026.

SAIA, Pedro Penalva; STRACHMAN, Eduardo. *In*: VI ENEI ENCONTRO NACIONAL DE ECONOMIA INDUSTRIAL, 2022, Salvador, BA. **Biotecnologia e Inovações Radicais: o Brasil entre 2009 e 2017**. Salvador, BA: [s. d.], 2022. p. 768–784.

SALVADOR-CARULLA, Luis *et al.* Adaptation of the technology readiness levels for impact assessment in implementation sciences: The TRL-IS checklist. **Heliyon**, [s. l.], v. 10, n. 9, p. e29930, 2024.

SAMARAH, Tariq *et al.* Intelligent Strategic Decision-Making for Optimized Project Management. *In*: 2024 INTERNATIONAL CONFERENCE ON DECISION AID SCIENCES AND APPLICATIONS (DASA), 2024, Manama, Bahrain. **2024 International Conference on Decision Aid Sciences and Applications (DASA)**. Manama, Bahrain: IEEE, 2024. p. 1–5. Disponível em: <https://ieeexplore.ieee.org/document/10836504/>. Acesso em: 21 jun. 2025.

SANTOS, Lariane Almeida *et al.* Metodologias de avaliação de investimento como suporte à decisão ex-ante e ex-post no desenvolvimento de patentes em icts públicas brasileiras: um estudo sobre a perspectiva da redução de risco de stranded patents. **Caderno Pedagógico**, [s. l.], v. 21, n. 4, p. e3838, 2024.

SANTOS, Bruna Gava Dos. **Modelagem De Corredores Ecológicos No Estado Do Paraná E Análise Comparativa Com as Áreas Estratégicas Para a Conservação E Restauração Da Biodiversidade**. 2024. 48 f. - UNIVERSIDADE ESTADUAL DO OESTE DO PARANÁ, [s. l.], 2024. Disponível em: Acesso em: 10 jun. 2025.

SATHUPADI, Kaushik *et al.* RAP-Optimizer: Resource-Aware Predictive Model for Cost Optimization of Cloud AlaaS Applications. **Electronics**, [s. l.], v. 13, n. 22, p. 4462, 2024.

SCOPUS. **Scopus - Document search | Signed in**. [S. l.], 2019. Disponível em: <https://www.scopus.com/search/form.uri?display=basic&zone=header&origin=searchbasic#basic>. Acesso em: 21 mar. 2025.

SHAW, Kassi; GOÏTA, Kalifa; GERMAIN, Mickaël. Prospectivity Mapping of Heavy Mineral Ore Deposits Based upon Machine-Learning Algorithms: Columbite-Tantalite Deposits in West- Central Côte d'Ivoire. **Minerals**, [s. l.], v. 12, n. 11, p. 1453, 2022.

SIEVERT, Carson. **Interactive Web-Based Data Visualization with R, plotly, and shiny**. 1. ed. [S. l.]: Chapman and Hall/CRC, 2020. Disponível em: <https://www.taylorfrancis.com/books/9780429824210>. Acesso em: 10 jun. 2025.

SILGE, Julia; ROBINSON, David. **Text mining with R: a tidy approach**. First edition. Beijing ; Boston: O'Reilly, 2017.

SILVA, Luan Carlos Santos *et al.* Assessment of Technology Transfer Processes at Brazilian Public Universities. **SAGE Open**, [s. l.], v. 15, n. 1, 2025. Disponível em: Acesso em: 1 jan. 2025.

SILVA, Alessandra Lopes da *et al.* Eficiência da metodologia da análise envoltória de dados para gestão organizacional. **Revista Universitária Brasileira**, [s. l.], v. 3, n. 1, p. 24, 2025.

SILVA, Luan Carlos Santos. Gestão da transferência de tecnologia para inovação nas universidades públicas brasileiras. **Revista Iberoamericana de Educación**, [s. l.], v. 95, n. 1, p. 17–40, 2024.

SILVA, Glênio Oliveira da; OLIVEIRA, Guilherme Saramago de; SILVA, Michele Maria da. Estudo De Caso Único: Uma Estratégia De Pesquisa. **Revista Prisma**, [s. l.], v. 2, n. 1, p. 78–90, 2021.

SIMAR, Léopold; WILSON, Paul W. Estimation and inference in two-stage, semi-parametric models of production processes. **Journal of Econometrics**, [s. l.], v. 136, n. 1, p. 31–64, 2007.

SOUZA, Marcel Mendes de. **Mapeamento Do Processo De Transferência De Tecnologia: Aplicação Na Área De Biotecnologia Da UFAM**. 2021. 1–198 f. PhD Thesis[s. l.], 2021. Disponível em: <https://tede.ufam.edu.br/handle/tede/8582>.

STAKOVIK, Flavio Henrique Moura. **Inovação em acesso à informação em biotecnologia: desenvolvimento de um mecanismo de busca inteligente baseado em processamento de linguagem natural**. 2023. 1–122 f. Tese - Universidade Católica de Brasília, Brasília - DF, 2023. Disponível em: <https://bdtd.ucb.br:8443/jspui/handle/tede/3357>.

STEHRMAN, Stephen V. Selecting and interpreting measures of thematic classification accuracy. **Remote Sensing of Environment**, [s. l.], v. 62, n. 1, p. 77–89, 1997.

TAMIRU, Habtamu; WAGARI, Meseret; TADESE, Bona. An Integrated Artificial Intelligence and GIS Spatial Analyst Tools for Delineation of Groundwater Potential Zones in Complex terrain: Fincha Catchment, Abay Basi, Ethiopia. **Air, Soil and Water Research**, [s. l.], v. 15, n. 15, p. 117862212110459–117862212110459, 2022.

TEIXEIRA, Miguel Paiva *et al.* A Aplicação Da Análise Envoltória De Dados Para Avaliação De Eficiência De Cursos De graduação: Uma Revisão sistemática/ the Application of Data Envelopment Analysis to Evaluate the Efficiency of Undergraduate courses: a Systematic Review. **Brazilian Journal of Development**, [s. l.], v. 7, n. 7, p. 69424–69443, 2021.

THANASSOULIS, Emmanuel; PORTELA, Maria C. S.; DESPIĆ, Ozren. Data Envelopment Analysis: The Mathematical Programming Approach to Efficiency Analysis. *In*: FRIED, Harold O.; LOVELL, C. A. Knox; SCHMIDT, Shelton S. (org.). **The Measurement of Productive Efficiency and Productivity Change**. [S. l.]: Oxford University Press, 2008. p. 251–420. Disponível em: <https://academic.oup.com/book/8538/chapter/154411312>. Acesso em: 17 maio 2026.

TONIAL, Graciele *et al.* CAPACIDADE ABSORTIVA E INOVAÇÃO: UM ESTUDO EM EMPRESAS DE BASE TECNOLÓGICA. *In*: CONGRESSO INTERNACIONAL DE CONHECIMENTO E INOVAÇÃO (CIKI), 2022. **Anais [...]**. [S. l.]: [s. d.], 2022. Disponível em: <https://proceeding.ciki.ufsc.br/index.php/ciki/article/view/1123>. Acesso em: 21 nov. 2025.

TONOLLI, Thaynara Gilli *et al.* Emprego Do Data Envelopment Analysis (dea) Para Estimar a Eficiência Relativa Dos Cursos De Pós-Graduação Da Ufsc No Período De Pandemia. [s. l.], p. 14, 2023a.

TONOLLI, Thaynara Gilli *et al.* Emprego Do Data Envelopment Analysis (dea) Para Estimar a Eficiência Relativa Dos Cursos De Pós-Graduação Da Ufsc No Período De Pandemia. *In*: DESAFIOS Y FUTURO DE LA EDUCACIÓN ANTE EL IMPACTO DE LA IA, 2023b, Assunção - Paraguai. **XXII Coloqui Internacional de Festión Universitaria**. Assunção - Paraguai: [s. d.], 2023. p. 1–14.

UFAM. **Plano de Desenvolvimento Institucional**. Manaus, AM: [s. d.], 2016. Relatório. Disponível em: <https://drive.google.com/file/d/18q2r5KP9NCnTb4luYTrHvpDjAe0CnIMf/view?pli=1>.

VALLE, Paulo Roberto Dalla; FERREIRA, Jacques De Lima. Análise De Conteúdo Na Perspectiva De Bardin: Contribuições E Limitações Para a Pesquisa Qualitativa Em Educação. **Educação Em Revista**, [s. l.], v. 41, 2025.

VELLANI JÚNIOR, Raymundo Lázaro *et al.* Analysis of the Evolution of the Number of Biotechnology Patents in the Agribusiness Sector. **Agriculture, Agribusiness and Biotechnology**, [s. l.], v. 65, p. 1–10, 2022.

VERMA, Pranay *et al.* ARON: Adaptive Resource Optimization Network for AI-Driven Business Management. *In*: 2024 4TH INTERNATIONAL CONFERENCE ON TECHNOLOGICAL ADVANCEMENTS IN COMPUTATIONAL SCIENCES (ICTACS), 2024, Tashkent, Uzbekistan. **2024 4th International Conference on Technological Advancements in Computational Sciences (ICTACS)**. Tashkent, Uzbekistan: IEEE, 2024. p. 61–66. Disponível em: <https://ieeexplore.ieee.org/document/10840557/>. Acesso em: 21 jun. 2025.

VIERA, Anthony J.; GARRETT, Joanne M. Understanding interobserver agreement: the kappa statistic. **Family Medicine**, [s. l.], v. 37, n. 5, p. 360–363, 2005.

VILHENA, Marcelo Azevedo. **Análise envoltória de dados aplicada ao diagnóstico do índice de inovação dos programas de pós-graduação da Universidade de Brasília**. 2025. 108 f. Dissertação - UnB, Brasília - DF, 2025. Disponível em: <http://repositorio.unb.br/handle/10482/52965>. Acesso em: 21 nov. 2025.

VILLELA, Gisele Karine Murador *et al.* Estudo Prospectivo de Patentes sobre Galactooligosacarídeos com Potencial Biotecnológico. **Ensaio e Ciência C Biológicas Agrárias e da Saúde**, [s. l.], v. 26, n. 2, p. 197–204, 2022.

WANG, Luoqi *et al.* Enhancing mineral prospectivity mapping with geospatial artificial intelligence: A geographically neural network-weighted logistic regression approach. **International Journal of Applied Earth Observation and Geoinformation**, [s. l.], v. 128, p. 103746–103746, 2024.

WANG, Yu-Hui; LIN, Guan-Yu. Exploring AI-healthcare innovation: natural language processing-based patents analysis for technology-driven roadmapping. **Kybernetes**, [s. l.], v. 52, n. 4, p. 1173–1189, 2023.

WEID, Irene von der *et al.* **Categorização de patentes de Biotecnologia baseada na Classificação Internacional de Patentes e análise do panorama de depósito de pedidos de patentes neste setor no Brasil (2012-2016)**. [S. l.]: INPI, 2018.

WILLMOTT, Cj; MATSUURA, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. **Climate Research**, [s. l.], v. 30, p. 79–82, 2005.

WIPO. **International Patent Classification (IPC)**. [S. l.], 1971. Disponível em: <https://www.wipo.int/web/classification-ipc>. Acesso em: 23 mar. 2025.

WONG, Tzu-Tsung; YEH, Po-Yang. Reliable Accuracy Estimates from k -Fold Cross Validation. **IEEE Transactions on Knowledge and Data Engineering**, [s. l.], v. 32, n. 8, p. 1586–1594, 2020.

XIONG, Xi; YANG, Guo-liang; GUAN, Zhong-cheng. Assessing R&D efficiency using a two-stage dynamic DEA model: A case study of research institutes in the Chinese Academy of Sciences. **Journal of Informetrics**, [s. l.], v. 12, n. 3, p. 784–805, 2018.

YESUFU, Lawal; ALAJLANI, Sami. Enhancing Strategic Decision-Making With AI: Unveiling the Untapped Potential. *In*: GÜNAR, Altuğ (org.). **Advances in Finance, Accounting, and Economics**. [S. l.]: IGI Global, 2025. p. 415–436. Disponível em: <https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/979-8-3693-7036-0.ch018>. Acesso em: 21 jun. 2025.

YIN, Robert K. **Case study research: design and methods**. 5. editioned. Los Angeles London New Delhi Singapore Washington, DC: SAGE, 2014.

YOO, Yongmin; LIM, Dongjin; KIM, Kyungsun. Artificial Intelligence Technology analysis using Artificial Intelligence patent through Deep Learning model and vector space model. [s. l.], p. 1–15, 2021a.

YOO, Yongmin; LIM, Dongjin; KIM, Kyungsun. **Patent Analysis Using Vector Space Model and Deep Learning Model: A Case of Artificial Intelligence Industry Technology**. [S. l.]: MATHEMATICS & COMPUTER SCIENCE, 2021b. Disponível em: <https://www.preprints.org/manuscript/202111.0208/v1>. Acesso em: 18 mar. 2025.

ZANCANARO, Márcio; MATHIAS, Thaís; HAHN, Ivanete Schneider. O papel das instituições de ensino superior na promoção da inovação e empreendedorismo para o

desenvolvimento regional. **Revista on Line De Política E Gestão Educacional**, [s. l.], v. 28, n. e023027, p. 20, 2024.

ZHANG, Gaomin *et al.* A hybrid approach combining data envelopment analysis and recurrent neural network for predicting the efficiency of research institutions. **Expert Systems with Applications**, [s. l.], v. 238, p. 122150, 2024.

ZHU, Xiaohui; HU, Chunhua; LUO, Ling. Research on Intelligent Portfolio Optimization Model Based on Big Data and Artificial Intelligence. *In*: 2024 INTERNATIONAL CONFERENCE ON INTERACTIVE INTELLIGENT SYSTEMS AND TECHNIQUES (IIST), 2024, Bhubaneswar, India. **2024 International Conference on Interactive Intelligent Systems and Techniques (IIST)**. Bhubaneswar, India: IEEE, 2024. p. 160–166. Disponível em: <https://ieeexplore.ieee.org/document/10734543/>. Acesso em: 21 jun. 2025.

ZÜLFİKAROĞLU, Sarp. AI-Driven Strategic Management and Decision Making for Energy Sector. **Next Frontier For Life Sciences and AI**, [s. l.], v. 8, n. 1, p. 91, 2024.

DMU_Avalia a	UFM G	UFPE L	US P	UNICAM P	UFFR R	UNES P	UFFB B	UFRG S	UF C	UFR J	UFBA A	UFFS J	UFAL L	UFPE E	UFSCA R	UFU U	UFG G	UN B	UFV V	UFG G	UFES S	UTFP R	UES C	UFM S	UFAM M	UFMA A	UFFA A	UFFR N	UFFS M
UFMG	1,00																												
UFPEL		1,00																											
USP			1,00																										
UNICAMP				1,00																									
UFFPR					1,00																								
UNESP						1,00																							
UFFPB							1,00																						
UFRGS		0,07	0,15	0,14	0,32							0,32																	
UFC		0,24	0,03	0,07		0,53						0,13																	
UFRJ			0,69																0,31										
UFBA				0,20		0,01	0,77												0,02										
UFFSJ												1,00																	
UFAL		0,05		0,02								0,88							0,05										
UFPE			0,16	0,26								0,24							0,27			0,08							
UFSCAR				0,26		0,14													0,60										
UFU		0,04		0,33			0,63																						
UFCG																	1,00												
UNB			0,11			0,16													0,72										
UFV																		1,00											
UFG				0,13		0,22						0,38							0,27										
UFES				0,26			0,38												0,37										
UTFFPR																					1,00								
UESC						0,12						0,46							0,27		0,15								
UFMS																							1,00						
UFAM				0,09			0,34												0,57										
UFMA						0,02						0,13																	
UFFPA			0,19			0,13						0,11						0,53		0,27		0,30							
UFFRN			0,10	0,38		0,01													0,27		0,24								
UFFSM																													1,00

C. MATRIZ DE FOLGAS (SLACKS)

Slacks_Output

DMU	SCOPUS	PUBMED	Patentes
UFMG			
UFPEL			
USP			
UNICAMP			
UFPR			
UNESP			
UFPB			
UFRGS		11,713	
UFC		63,878	
UFRJ		55,699	4,291
UFBA			
UFSJ			
UFAL		1,169	
UFPE		11,164	
UFSCAR		19,373	
UFU		6,132	
UFCG			
UNB			12,402
UFV			
UFG		31,296	
UFES		2,807	
UTFPR			
UESC		3,153	
UFMS			
UFAM		3,109	
UFMA	0,876		11,125
UFPA			12,333
UFRN			34,959
UFSM			

Slacks_Input

DMU	Grad_Biotec	Bols_CAPES	Proj_CAPES
UFMG			
UFPEL			
USP			
UNICAMP			
UFPR			
UNESP			
UFPB			
UFRGS			
UFC			
UFRJ	218,375		555,697
UFBA	177,729		830,866
UFSJ			
UFAL			171,591
UFPE			
UFSCAR	144,854		683,252
UFU	92,576		94,173
UFCG			
UNB	132,544		653,408
UFV			
UFG			235,678
UFES	399,244		58,986
UTFPR			
UESC			127,540
UFMS			
UFAM	46,186		442,400
UFMA			
UFPA			
UFRN			
UFSM			

ANEXO 2

CAPÍTULO 1 - COMPROVANTE DE PUBLICAÇÃO



Transferência tecnológica: fatos e percepções sobre patentes em biotecnologia no Brasil de 2014 a 2023

Technology transfer: facts and perceptions about biotechnology patents in Brazil from 2014 to 2023

Transferencia de tecnología: datos y percepciones sobre las patentes en biotecnología en Brasil de 2014 a 2023

DOI: 10.55905/oelv23n2-072

Receipt of originals: 1/17/2025

Acceptance for publication: 2/10/2025

Wilson Kume

Doutorando em Gestão da Inovação em Biotecnologia

Instituição: Universidade Federal do Amazonas

Endereço: Manaus, Amazonas, Brasil

E-mail: wilson.kume@ufam.edu.br

Manoel Carlos de Oliveira Junior

Doutor em Biotecnologia

Instituição: Universidade Federal do Amazonas

Endereço: Manaus, Amazonas, Brasil

E-mail: manoelcarlos@ufam.edu.br

RESUMO

Este estudo trata da evolução das patentes no setor de biotecnologia entre 2014 e 2023 no Brasil e seu papel na promoção de inovações capazes de transformar setores como saúde, agricultura, energia, meio ambiente e indústria. A biotecnologia é a união do conhecimento científico com o tecnológico, proporcionando soluções para desafios mundiais, tais como a segurança alimentar e o tratamento das doenças. Apesar dos avanços, dificuldades ainda são enfrentadas, incluindo a dependência das inovações estrangeiras e a baixa taxa de patentes na indústria biotecnológica. O estudo enfatiza a importância de um ambiente que vise proteger as invenções, mas também a promoção da pesquisa e do desenvolvimento. A colaboração entre universidades e empresas é necessária para o aprimoramento da inovação, porém há barreiras a serem superadas, como a burocracia e a falta de dados integrados. Conclui-se que, para revitalizar o setor, é necessário um esforço conjunto que estimule a troca de conhecimentos e um sistema de patentes mais eficiente para possibilitar que o país tenha melhor desempenho em um mundo competitivo.

Palavras-chave: Patentes, Biotecnologia, Inovação, Cooperação.

ANEXO 3

CAPÍTULO 2 - COMPROVANTE DE PUBLICAÇÃO



DOI: 10.53660/PRW-2922-5210

Eficiência no desempenho inovador em biotecnologia: uma análise envoltória de dados (DEA) nas universidades públicas brasileiras

Efficiency in innovative performance in biotechnology: a data envelopment analysis (DEA) in brazilian public universities

Received: 15/04/2026 | Accepted: 10/05/2026 | Published: 15/05/2026

Wilson Kume

Universidade Federal do Amazonas, Brasil
ORCID: <https://orcid.org/0000-0003-1542-7720>
E-mail: wilson.kume@ufam.edu.br

Maria do Perpétuo Socorro Rodrigues Chaves

Universidade Federal do Amazonas, Brasil
ORCID ID: 0000-0003-4289-2257
E-mail: socorro.chaves@ufam.edu.br

Manoel Carlos de Oliveira Júnior

Universidade Federal do Amazonas, Brasil
ORCID ID: 0000-0003-4630-0810
E-mail: manoelcarlos@ufam.edu.br

RESUMO

A presente pesquisa se dedica a examinar como as universidades públicas do Brasil aproveitam seus recursos acadêmicos para gerar inovação no promissor campo da biotecnologia. A investigação recorre à Análise Envoltória de Dados (DEA) em dois níveis para avaliar a eficiência produtiva das instituições, quantificando de que maneira o suporte financeiro (bolsas e projetos) e o capital humano (formados) se traduzem em ativos de Propriedade Intelectual (PI) e publicações científicas. Diferente de abordagens que focam apenas em contratos de licenciamento, muitas vezes opacos nos registros oficiais, este trabalho estabelece uma fronteira de eficiência baseada na geração de patentes 'Biotec BR'. Os resultados permitem identificar *benchmarks* de produtividade tecnológica e propõem um modelo de super-eficiência que integra indicadores de inovação qualitativos de impacto institucional. A análise contribui para o entendimento da lacuna entre a produção científica e a prontidão tecnológica no setor biotecnológico nacional.

Palavras-chave: Transferência de Tecnologia; Biotecnologia; Análise Envoltória de Dados (DEA); Eficiência.

ANEXO 4

CAPÍTULO 3 - COMPROVANTE DE PUBLICAÇÃO

REVISTA OWL (*OWL Journal*)

www.revistaowl.com.br – ISSN: 2965-2634



POTENCIAL DE PATENTEAMENTO E MATURIDADE TECNOLÓGICA: ANÁLISE INTEGRADA COM INTELIGÊNCIA ARTIFICIAL NO PPGBIOTEC/UFAM

*PATENTING POTENTIAL AND TECHNOLOGICAL MATURITY:
INTEGRATED ARTIFICIAL INTELLIGENCE ANALYSIS AT
PPGBIOTEC/UFAM*

*POTENCIAL DE PATENTAMIENTO Y MADUREZ TECNOLÓGICA: UN
ANÁLISIS INTEGRADO CON INTELIGENCIA ARTIFICIAL EN
PPGBIOTEC/UFAM*

DOI: 10.5281/zenodo.20114796



Wilson Kume¹

Maria do Perpétuo Socorro Rodrigues Chaves²

Manoel Carlos de Oliveira Júnior³

RESUMO

Este trabalho efetiva uma análise com Inteligência Artificial no Ecossistema de Biotecnologia do no Programa de Pós-Graduação em Biotecnologia da Universidade Federal do Amazonas–PPGBIOTEC/UFAM. No estudo foram avaliadas, em relação à sua maturidade tecnológica e ao potencial de se tornarem patentes, as dissertações e teses defendidas entre 2015 e 2024 no PPGBIOTEC/UFAM. Técnicas de Processamento de Linguagem Natural foram utilizadas para identificar termos associados aos códigos da Classificação Internacional de Patentes (CIP) do modelo

1 Doutorando em Gestão da Inovação em Biotecnologia PPGBIOTEC/UFAM - Universidade Federal do Amazonas - AM, Brasil. E-mail: wilson.kume@ufam.edu.br

2 Professora Doutora PPGBIOTEC/UFAM - Universidade Federal do Amazonas - AM, Brasil. E-mail: socorro.chaves@ufam.edu.br

3 Professor Doutor FES/UFAM - Universidade Federal do Amazonas - AM, Brasil. E-mail: manoelcarlos@ufam.edu.br

Revista *OWL Journal*, Campina Grande – PB, v.4 n.5 (2026) – ISSN 2965-2634

A Revista *OWL Journal* está licenciada com uma Licença Creative Commons Atribuição

(CC BY)

1/24



ANEXO 5

CAPÍTULO 4 - COMPROVANTE DE PUBLICAÇÃO



OTIMIZAÇÃO DA BIOPROSPECÇÃO DE ATIVOS BIOLÓGICOS POR INTELIGÊNCIA ARTIFICIAL

OPTIMIZATION OF BIOPROSPECTING OF BIOLOGICAL ASSETS BY ARTIFICIAL INTELLIGENCE

OPTIMIZACIÓN DE LA BIOPROSPECCIÓN DE ACTIVOS BIOLÓGICOS MEDIANTE INTELIGENCIA ARTIFICIAL

 10.56238/revgeov17n4-029

Wilson Kume

Doutorando em Gestão da Inovação em Biotecnologia - PPGBIOTEC

Instituição: Universidade Federal do Amazonas

E-mail: wilson.kume@ufam.edu.br

Orcid: <https://orcid.org/0000-0003-1542-7720>

Junior de Jesus Costa

Doutorando em Gestão da Inovação em Biotecnologia – PPGBIOTEC

Instituição: Universidade Federal do Amazonas

E-mail: costa.junior@ufam.edu.br

Orcid: <https://orcid.org/0000-0003-4540-3375>

RESUMO

O presente estudo examina a identificação de ativos biológicos por meio da união estratégica entre dados ambientais georreferenciados e modelos preditivos, com base em inteligência artificial (IA). Seu objetivo principal é a otimização dos processos e a diminuição das despesas operacionais na bioprospecção tradicional. Para tal, parte-se do pressuposto de que a biodiversidade terrestre é um vetor de ativo de valor incalculável, evidenciando a urgência em se criar e operacionalizar metodologias úteis para lidar com essa complexidade, transpondo as limitações das metodologias tradicionais. Desse modo, aplicou-se o algoritmo *Random Forest* junto a *rasters* climáticos, *shapefiles* e dados do Sistema de Informação sobre a Biodiversidade Brasileira (SiBBR), visando mapear áreas de elevado potencial bioativo. A metodologia foi testada por meio de protótipo em R Shiny, que possibilitou simular cenários pragmáticos e mapear riscos operativos. Mesmo os resultados preliminares apontando para a necessidade de refino paramétrico, os achados ratificam o potencial das ferramentas de IA em servir como agentes catalisadores à bioeconomia sustentável na Amazônia.

Palavras-chave: Bioprospecção. Inteligência Artificial. *Random Forest*. Otimização de Prospeção.

ABSTRACT

The present study examines the identification of biological assets through the strategic union between georeferenced environmental data and predictive models, based on artificial intelligence (AI). Its main objective is the optimization of processes and the reduction of operating expenses in traditional bioprospecting. To this end, it assumes that terrestrial biodiversity is an asset vector of incalculable value, evidencing the urgency of creating and operationalizing useful methodologies to deal with this

