

Estado da publicação: O preprint não foi publicado em outro meio.

Um protocolo para aproveitamento de IA no GEE para mapeamento de transição de uso e cobertura da terra: o caso de São Paulo (1985-2024)

Rodrigo da Cunha Pacheco

<https://doi.org/10.1590/SciELOPreprints.13795>

Submetido em: 2025-10-16

Postado em: 2025-10-23 (versão 1)

(AAAA-MM-DD)

Um protocolo para aproveitamento de IA no GEE para mapeamento de transição de uso e cobertura da terra: o caso de São Paulo (1985-2024)

Dr. Rodrigo da Cunha Pacheco

Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP), Campus Jacareí.

ORCID: <https://orcid.org/0000-0001-5497-9793>

E-mail: rodrigo.pacheco@ifsp.edu.br

RESUMO

Este artigo explora a aplicação de inteligência artificial generativa como um assistente de codificação para produzir mapas de transição de uso e cobertura da terra dentro da plataforma Google Earth Engine (GEE). Focando na expansão urbana no Estado de São Paulo de 1985 a 2024, propõe-se um protocolo que permite aos pesquisadores — particularmente aqueles com limitada expertise em programação — aproveitar o poder da IA para utilizar as extensas capacidades de geoprocessamento do GEE. O protocolo implementa concorrentemente camadas de garantia de qualidade para assegurar a acurácia do mapa final. Os resultados bem-sucedidos do mapeamento demonstram a viabilidade e a eficácia desta abordagem.

Palavras-chave: Geoprocessamento; Inteligência Artificial; Google Earth Engine; Expansão Urbana.

A protocol for leveraging AI in GEE for land use and land cover transition mapping: the Case of São Paulo (1985-2024)

ABSTRACT

This paper explores the application of generative artificial intelligence as a coding assistant to produce land use and land cover transition maps within the Google Earth Engine (GEE) platform. Focusing on urban expansion in the State of São Paulo from 1985 to 2024, we

propose a protocol that enables researchers—particularly those with limited programming expertise—to harness the power of AI to utilize GEE's extensive geoprocessing capabilities. The protocol concurrently implements quality assurance layers to guarantee the accuracy of the final map product. The successful results of the mapping demonstrate the viability and effectiveness of this approach.

Keywords: Geoprocessing; Artificial Intelligence; Google Earth Engine; Urban Expansion.

Un protocolo para el aprovechamiento de la IA en GEE para el mapeo de la transición de uso y cobertura del suelo: el caso de São Paulo (1985-2024)

RESUMEN

Este artículo explora la aplicación de la inteligencia artificial generativa como asistente de codificación para producir mapas de transición de uso y cobertura del suelo dentro de la plataforma Google Earth Engine (GEE). Centrándose en la expansión urbana en el Estado de São Paulo de 1985 a 2024, se propone un protocolo que permite a los investigadores — particularmente a aquellos con experiencia limitada en programación— aprovechar el poder de la IA para utilizar las extensas capacidades de geoprocesamiento de GEE. El protocolo implementa simultáneamente capas de garantía de calidad para asegurar la precisión del mapa final. Los exitosos resultados de la cartografía demuestran la viabilidad y eficacia de este enfoque.

Palabras clave: Geoprocesamiento; Inteligencia Artificial; Google Earth Engine; Expansión Urbana.

1. Introdução

O crescente desenvolvimento tecnológico alcançou um novo patamar em que a inteligência e o raciocínio humanos, elementos centrais na resolução de problemas complexos, passaram a ser simulados ou similarmente desempenhados por meio de máquinas. Como toda novidade tecnológica, mas em maior intensidade, a Inteligência Artificial (IA), especialmente os *LLMs* (Grande Modelo de Linguagem), tem gerado uma diversidade de reações: pessimistas e otimistas. Fato é que seu uso pode ter implicações boas ou ruins (Ludermir, 2021), motivo pelo qual tem demandado muita discussão. No

campo científico, a visão conservadora é que a IA generativa possui equivalência paradigmática à criação do primeiro periódico científico em 1665. Mas o impacto pode ser bem maior, promovendo a completa reformatação da cultura e das práticas científicas (Vasconcelos; Marušić, 2025).

No ambiente acadêmico do geoprocessamento e da análise espacial, as implicações da IA são particularmente promissoras. Pesquisadores têm defendido que o desenvolvimento de modelos como o *LLM* pode trazer grandes benefícios. O grande destaque é a habilidade dessa tecnologia em relação a fatores cruciais: compreensão da linguagem natural, raciocínio sofisticado e, notavelmente, geração de códigos. Essas capacidades podem ser usadas para automatizar as atividades do geoprocessamento, ainda que parcialmente. Vislumbra-se a era dos Sistemas de Informação Geográfica (SIGs) autônomos, capazes de desempenhar atividades de alta complexidade (Akinboyewa et al., 2025; Li et al., 2025).

As expectativas não devem ignorar, contudo, os desafios associados à ferramenta. Preocupações relacionadas à confiabilidade dos processos, à acurácia dos produtos gerados e às implicações éticas devem ser igualmente abordadas para a manutenção da integridade científica, a qual “está atrelada à adesão a práticas científicas éticas, transparentes e rigorosas” (Vasconcelos; Marušić, 2025).

Nesse contexto, o presente estudo busca contribuir com a formulação de protocolos que permitam aos pesquisadores, mesmo com pouca noção de codificação, utilizar o geoprocessamento em nuvem por meio do Google Earth Engine (GEE), aproveitando-se de seu grande poder computacional. Assim, recorre-se à capacidade de geração de código das IAs generativas, a partir de instruções em linguagem natural, para tornar o GEE mais acessível. O protocolo de segurança aqui proposto, no entanto, visa garantir que o produto resultante possua o nível de confiança exigido pela integridade científica.

Para validar o protocolo, sua aplicação foi realizada a partir de um estudo de caso. Trata-se da produção de um mapa de transição de uso e cobertura da terra focado na expansão urbana do Estado de São Paulo entre 1985 e 2024. É importante ressaltar que o protocolo proposto é de natureza específica, sendo, neste primeiro momento, projetado para a geração de mapas que abordem a transição a partir de fontes confiáveis já existentes em formato raster, como as coleções do MapBiomias.

Ao fazer isso, o estudo visa trazer para o debate as oportunidades e os desafios que o uso da IA generativa, em conjunto com as geotecnologias, pode promover em qualquer atividade que envolva a informação e o raciocínio geográfico. Espera-se, assim, contribuir para a análise técnica e metodológica da produção confiável de informações geográficas na era da IA generativa.

2. Fundamentos teóricos e problematizações

Duas grandes transformações possibilitaram avanços significativos nas pesquisas geográficas contemporâneas. De um lado, a evolução das teorias e metodologias, que permitiram abordar uma multiplicidade de temas de forma sistêmica, e não apenas enciclopédica. De outro, o avanço das geotecnologias, que reforçou a abordagem sintética por meio de técnicas que viabilizam a representação e a modelagem mais complexa dos fenômenos espaciais, com maior volume e velocidade no tratamento de dados. Compreendem as geotecnologias todos os instrumentos que permitem coletar e processar informações geograficamente referenciadas (Rosa, 2011).

A crescente simbiose entre a Geografia e os aparatos tecnológicos fez emergir conceitos adicionais como Geografia Tecnológica e Geografia Automatizada (Fitz, 2008). Enquanto isso, a revolução representada pela *Web 2.0* trouxe também novas possibilidades de interação entre o usuário e os sistemas na nuvem. Emergiu então mais um conceito: a Geografia 2.0, que ocorre por meio da *Web 2.0* e representa a possibilidade de interatividade entre os usuários da informação geográfica e os sistemas baseados na nuvem. Trata-se da Geografia Interativa acessada e produzida por meio da nuvem (Morgato, 2011).

A relevância socioeconômica da Geografia 2.0 é evidenciada quando se depara com o fato de as chamadas *Big Techs* investirem quantias significativas na criação de infraestruturas, produtos e serviços diretamente relacionados à geoinformação. Muitos deles, como o Google Maps, o Google Earth e o Bing Maps, são de fácil utilização e acesso, tornando-se populares. Para usuários mais especializados e com demandas mais sofisticadas relacionadas à geoinformação, existe o Google Earth Engine (GEE). Trata-se de: “[...] uma plataforma baseada em nuvem para análise geoespacial em escala planetária que utiliza as imensas capacidades de computação do Google para abordar uma variedade de questões sociais de grande impacto [...]” (Gorelick *et al.*, 2017, tradução própria). Segundo Yang *et al.* (2022, p. 1, tradução própria), “O GEE também oferece acesso à vasta maioria de dados de sensoriamento remoto multitemporal, públicos e disponíveis gratuitamente”.

Trabalhar com dados raster de grandes áreas, como um estado, um país ou mesmo continentes e o mundo, seria um desafio para o pesquisador comum, sem o acesso a caros sistemas informacionais, devido à elevada capacidade computacional requerida. Mas, se por um lado o GEE pode representar uma solução para a necessidade de poder de processamento, por outro, pode inibir seu uso por parte de usuários com pouca ou nenhuma familiaridade com linguagens de programação, como o JavaScript e Python (as linguagens de programação para a plataforma).

A solução de acessibilidade que surge é exatamente a IA generativa e a sua capacidade de raciocinar, codificar e compreender a linguagem natural. Ela é uma

tecnologia que permite a interação de usuários, leigos em tecnologia da informação e programação, com ferramentas inteligentes de propósitos gerais e mesmo especializadas em codificação e desenvolvimento de software (Krakowski, 2025). Isso se deve ao fato de os *LLMs* estarem aumentando rapidamente a sua acurácia. Alguns *benchmarks* mostram que o Gemini aumentou significativamente sua capacidade de programação da sua versão 1.5 Pro para 2.5 Pro (de 30,5% para 74,2% no LiveCodeBench e de 16,9% para 82,2% no Aider Polyglot) (Gemini Team, 2025).

Conforme Russel e Norvig (2013), não há definição unívoca para a IA. No contexto dos *Chatbots* atuais, baseados em *LLM*, é lógico caracterizá-los como ferramentas que executam funções que, nos seres humanos, exigiriam o uso da inteligência ou do raciocínio. Trata-se de uma definição suficientemente informativa do ponto de vista cognitivo e neutra do ponto de vista ontológico, adequada para os propósitos deste trabalho. Ocorre que há uma disputa entre as teses da IA fraca e forte: a primeira apenas assume a similitude das ações da IA com a inteligência humana, a segunda admite a possibilidade de uma IA realmente pensante e, eventualmente, consciente. A definição acima se alinha momentaneamente com a tese fraca mais conservadora.

O primeiro trabalho relacionado à IA surgiu em 1943, mas foi a partir de 2020 que o panorama se alterou radicalmente com o desenvolvimento dos *LLMs* para guiar os *chatbots*. Isso ocorreu devido às pesquisas acumuladas e pela disponibilização inédita de vastos conjuntos de dados para treinamento dos modelos. Por meio da linguagem natural, o usuário comum passou a ter a possibilidade de estabelecer tarefas para execução autônoma da IA. Os horizontes de aplicação tornaram-se vastos: educação, pesquisa, saúde, negócios, engenharia de software, etc. (Dam *et al.* 2024).

Ironicamente, apesar dos grandes e contínuos aprimoramentos, a IA generativa também se assemelha aos humanos na produção de falhas. As principais decorrem da desatualização da base de treinamento e das alucinações: quando ela fornece ou respostas comprovadamente falsas, ou que não podem ser confirmadas, ou ainda, que não condizem com as entradas (Dam *et al.*, 2024; Zhang *et al.*, 2025). Há também a questão da correta apreensão das intencionalidades do usuário a qual pode sofrer interferência da própria qualidade dos *prompts* de entrada. Isso significa que seu uso não pode ser feito de forma completamente livre de riscos da produção de erros.

A dupla situação que se apresenta, portanto, é o fato de a IA generativa ser potencialmente capaz de intermediar a programação das atividades com o GEE, ao converter as intenções do pesquisador em códigos de Javascript, ao mesmo tempo que não é capaz de o fazer de forma completamente livre da possibilidade de erros, seja na codificação ou mesmo na correta compreensão das intencionalidades do usuário. Dessa

forma, a busca por uma conciliação do benefício de acessibilidade ao GEE que a IA pode promover com a necessidade de preservação da integridade científica se impõe.

2.1. O protocolo de segurança

O fato de a IA não estar imune a alucinações não é um argumento suficiente para repelir completamente o seu uso na produção de *scripts* para o GEE. Mesmo no contexto da Engenharia de Software existem inúmeras técnicas desenvolvidas para captar possíveis erros produzidos a partir da codificação humana. Pfleeger (2004) considerou apenas como ideal o cenário no qual os programas elaborados sempre funcionassem ao serem executados, devendo por isso ocorrer o teste após a codificação. Por meio do teste, o software se torna confiável (Peters, Pedrycz, 2001). Assim, se nem o uso da IA e nem a programação completamente humana fornecem garantia, por si, da ausência de possíveis erros, a abordagem mais adequada em qualquer um dos casos é adotar procedimentos de minimização de riscos.

Dado esse contexto, é conveniente mencionar um exemplo de inspiração: no domínio da segurança cibernética, emprega-se a técnica de defesa em profundidade para minimizar riscos de falha. Trata-se da criação de uma série de camadas de proteção ou segurança que no conjunto visam tornar tais riscos cada vez mais improváveis. É uma adaptação proveniente das estratégias militares (Silva, 2022). O protocolo aqui proposto se utiliza exatamente do princípio de defesa em camadas contra eventuais erros de codificação por parte da IA generativa. Ressalta-se: ele está projetado, em um primeiro momento, para tornar confiável a geração de mapas de transição do uso e cobertura da terra, sendo caracterizado a seguir.

A primeira camada de segurança, após a obtenção do *script* na interação com a IA, é simplesmente rodá-lo no próprio GEE, pois erros que inviabilizam a execução do código serão apontados pelo próprio *console*, ajudando decisivamente na depuração do código. No estudo de caso que será apresentado, foi necessário realizar algumas vezes a depuração do código por erros cometidos pela IA. A mensagem de erro do *console* era fornecida para a IA que então buscava corrigi-lo. Por vezes, o processo de depuração teve que ser feito em sequência até que o código pudesse, finalmente, gerar um produto. Trata-se, então, de um primeiro processo superável com base na estratégia de tentativa e erro.

Muito embora o erro que não permite a execução completa do código seja frustrante, ele é menos perigoso que um erro que não inviabiliza a execução, mas gera um produto incorreto. Tais erros podem ser de inúmeros graus de gravidade, desde os mais grosseiros, os quais uma simples inspeção visual no produto pode constatar, até aqueles mais sutis e

arriscados. De toda forma, a fonte desses erros será ou a lógica errônea do *script* gerado pela IA, ou o uso de funções erradas do GEE.

Nesse contexto, a segunda camada de proteção é a inspeção visual e qualitativa do produto final obtido, a fim de verificar possíveis erros grosseiros a partir do conhecimento de domínio do pesquisador a respeito das características esperadas para a área de estudo como um todo ou de zonas específicas. Assim, por exemplo, se um pesquisador sabe que uma zona deveria ser de cobertura vegetal e o produto gerou uma área urbana, então, a presença de algum erro de *script* é inevitável. Para o caso de um estudo de uma área desconhecida, o pesquisador pode fazer um breve levantamento, por meio de imagens históricas, como no Google Earth, a respeito do conteúdo esperado para algumas zonas a serem comparadas com as obtidas pelo produto gerado no GEE. Trata-se da adaptação de uma metodologia de verificação do código passível de ser classificada como o que em Engenharia de Software é conhecido como teste de caixa-fechada. Conforme Pflieger (2004), neste tipo de teste, o conteúdo da caixa (o código) permanece desconhecido. O teste se concentra em fornecer a entrada (neste caso, o raster com os dados de cobertura da terra) e observa se as saídas produzidas são as esperadas (se algumas zonas de transição são realmente da classe que se esperava).

A terceira camada de segurança busca quantificar a acurácia da saída, ainda em um modelo de teste de caixa fechada, porém, mais rigoroso e capaz de capturar erros mais sutis, os quais uma primeira análise visual poderia deixar passar. Trata-se de uma adaptação da técnica conhecida em Sensoriamento Remoto como a Verdade Terrestre, na qual o pesquisador seleciona amostras no terreno ou de fonte confiável e as compara com os *pixels* e suas respectivas classes de classificação em um mapa de cobertura da terra gerado a partir de classificação supervisionada ou não supervisionada. O procedimento visa construir uma matriz de confusão na qual a diagonal mostrará quantos pontos de amostras foram classificados corretamente, permitindo gerar índices de acurácia (Souza, 2020).

Assim, o protocolo estabelece um método de validação, denominado “Verdade de Referência”, focado em garantir a acurácia do mapa de transição em relação à combinação das imagens de entrada — e não em atestar a qualidade dessas imagens, como faria a Verdade Terrestre. Para isso, no desenho do *script* de auditoria, a IA é instruída a implementar uma lógica de amostragem estratificada sobre as imagens originais. O *script* deve coletar *pixels* de amostra, aplicar uma lógica condicional para prever a classificação correta no mapa de transição e, por fim, comparar esses valores esperados com os resultados gerados pelo *script* principal no mapa de transição. Com base nessa comparação, uma matriz de confusão é construída para extrair as estatísticas de acurácia.

O subprotocolo de verificação e validação do *script* principal por meio da Verdade de Referência é implementado por um *script* auxiliar de auditoria, gerado também com a

assistência da IA. Observa-se que processos de testes automatizados não são estranhos às práticas usadas em Engenharia de Software, como ocorre na análise estática automatizada (Sommerville, 2007). Contudo, para garantir uma independência genuína e mitigar o risco de falhas de causas comuns ou sistêmicas (em que vieses do modelo poderiam levar a erros idênticos), a abordagem separa a concepção da lógica do algoritmo da sua implementação em códigos. Assim, o pesquisador desenha o fluxo lógico e as etapas processuais da verificação, delegando à IA apenas a tarefa de traduzir este desenho em código funcional.

Com isso, criam-se dois caminhos computacionais para o mesmo resultado pretendido, cujas lógicas de origem são cognitivamente distintas: uma autônoma da IA (*script* principal) e outra humana (*script* auditor). Embora a IA, como implementadora, ainda possa introduzir erros de codificação no *script* auditor, torna-se improvável que um erro de lógica no *script* principal convirja, por acaso, para o mesmo resultado de um erro distinto de implementação no *script* auditor. Portanto, a eventual concordância entre os resultados produzidos pelos dois *scripts* oferece um forte indício de que a lógica autônoma da IA está correta e que ambas as implementações são funcionalmente confiáveis.

A quarta camada de proteção constitui o que pode ser denominado de teste da Caixa Aberta Suave. Em Engenharia de Software os testes de caixa aberta se voltam para o código e sua lógica (Oliveira, 2015). O fator limitante, contudo, é o possível pouco conhecimento em codificação do usuário para o qual este protocolo foi especialmente pensado. Embora o usuário não possa atestar a qualidade da sintaxe do código, pode pedir para a IA explicar detalhadamente cada linha e/ou etapa, tornando possível, para ele, auditar pelo menos a lógica implementada. É um procedimento que estabelece igualmente uma curva de aprendizado em relação ao processo de codificação. Novamente, há o risco de alucinações da IA, mas a probabilidade de erros completos para modelos que já estão avançados é baixa e eventuais erros mais pontuais poderão ser minimizados pela ação conjunta das quatro camadas de proteção desenvolvidas.

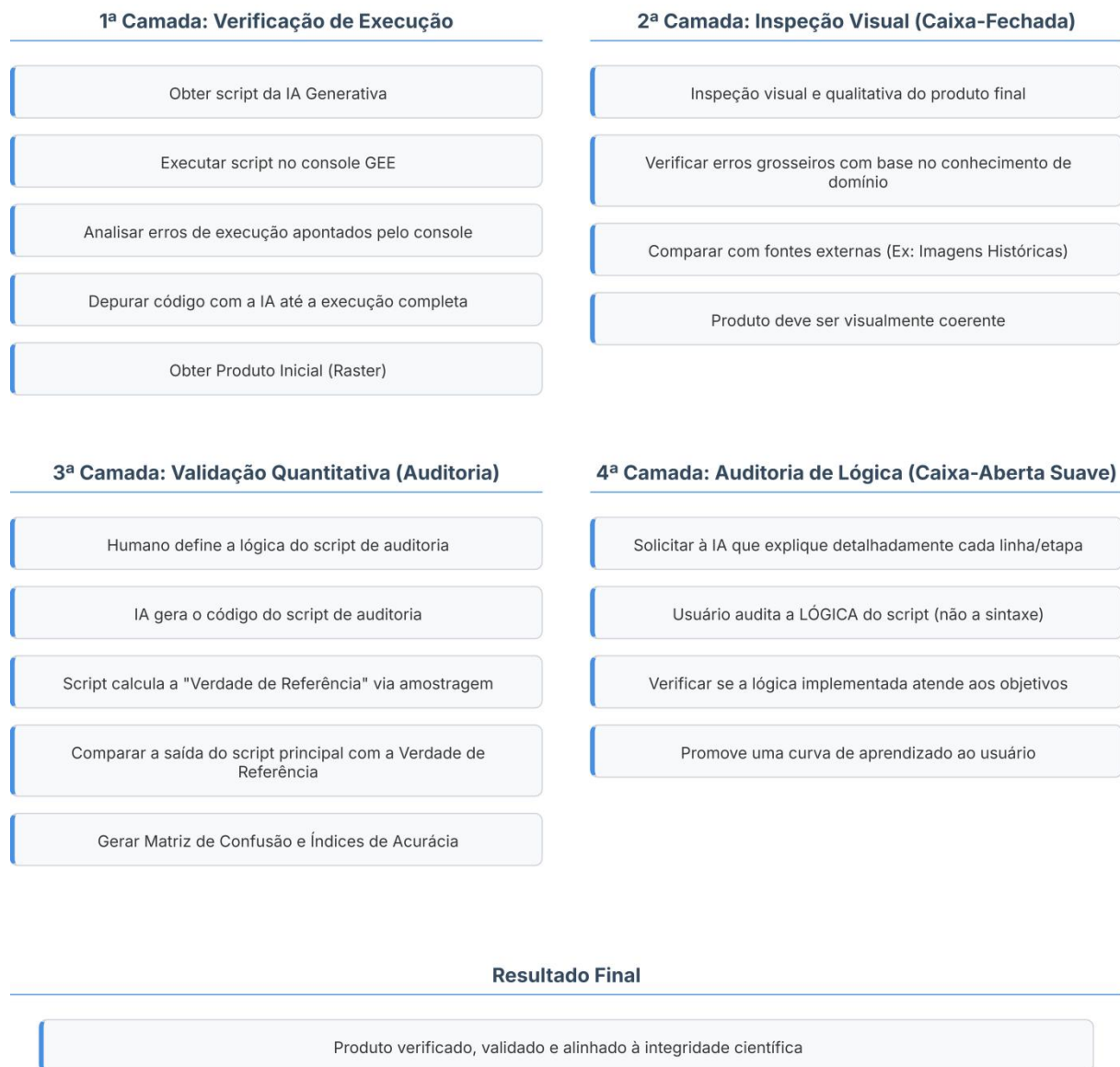
A Figura 1 ilustra as camadas de segurança do sistema de proteção em profundidade. A detecção de falhas insolúveis (o que não ocorreu no estudo de caso) poderá remeter o processo para um novo início.

Figura 1. Camadas do sistema de proteção em profundidade.



O protocolo proposto permite ao pesquisador verificar e validar os códigos gerados com o auxílio da IA. Em Engenharia de Software, a verificação busca demonstrar o alinhamento de um programa às suas especificações (ao que deve ser capaz de fazer), enquanto a validação busca verificar se o programa atende às necessidades do usuário (Sommerville, 2007). Ao final, o produto obtido estará alinhado à necessidade de integridade científica, com riscos de qualidade bastante reduzidos. A Figura 2 apresenta um diagrama com os pontos principais do protocolo e sua síntese.

Figura 2. Características principais do protocolo de defesa em profundidade.



3. Materiais e Métodos

Busca-se agora apresentar, de forma bastante detalhada, os materiais e procedimentos metodológicos mais relevantes da pesquisa, os quais implementam o protocolo de verificação e validação do código gerado pela IA para o GEE.

3.1. Materiais e *Software*

Para aplicação do protocolo, foram utilizados os seguintes dados e plataformas:

3.1.1. Dados Vetoriais:

3.1.1.1. Arquivo (*shapefile*) da malha de unidades federativas do Brasil, obtido do Instituto Brasileiro de Geografia e Estatística (IBGE, 2025a).

3.1.1.2. Arquivo (*shapefile*) das rodovias do Estado de São Paulo (São Paulo, 2025).

3.1.2. Dados Raster originais:

3.1.2.1. Mapeamentos anuais da Coleção 10 do MapBiomas Brasil (2025a), com resolução espacial de 30 metros, abrangendo o período de 1985 a 2024.

3.1.3. Plataformas e *Softwares*:

3.1.3.1. Google Earth Engine (GEE): Plataforma de geoprocessamento em nuvem utilizada para geração do arquivo raster de transição do uso e cobertura da terra.

3.1.3.2. Gemini Pro 2.5: Usado para criar o *script* em JavaScript para geração do raster de transição no GEE, das estatísticas e também para auxiliar no processo de verificação e validação do código.

3.1.3.3. QGIS 3.34.11: Software livre de Sistema de Informação Geográfica (SIG) empregado para a manipulação do raster de saída do GEE, análises visuais e elaboração do produto cartográfico final.

3.2. Procedimentos Metodológicos

O fluxo de trabalho foi executado em 8 etapas principais, sete das quais pertencentes e sintetizadas no diagrama de defesa em profundidade (Figura 2).

3.2.1. Fase 1: Inicialmente foram fornecidos os códigos de uso e cobertura da terra da coleção 10 do MapBiomas para o Gemini. Também foram fornecidos os *assets* (endereços internos do GEE) do limite do Estado de São Paulo e da própria coleção 10.

3.2.2. Fase 2: Pediu-se então para a IA criar um mapa de transição de uso e cobertura da terra usando os arquivos raster de 1985 e 2024, tendo como área de interesse o Estado de São Paulo. Das inúmeras classes de cobertura da terra do MapBiomas, foi determinado que o mapa de transição deveria ter somente três classes: área urbana estável, urbanização e outras (para todas as outras situações). Deixou-se claro que o objetivo principal era mapear o aumento da

mancha urbana no período. Foi determinado também o cálculo das estatísticas do tamanho da área ocupada por cada nova classe em km², hectare e percentual em relação ao tamanho do estado. A metodologia de formação do mapa de transição não foi especificada, assim, coube à IA implementar uma solução. Como o *script* inicial não rodou completamente por erros, forneceu-se os erros apresentados no console do GEE para o próprio Gemini resolver (camada 1 do protocolo de segurança), até um raster de saída ter sido gerado. Para identificar a dinâmica da urbanização, a IA aplicou uma técnica de álgebra de mapas sobre os arquivos raster originais, gerando mapeamentos binários com os seguintes valores possíveis para os *pixels*: 1, representando a classe "Infraestrutura Urbana" (código 24 da legenda MapBiomias), e 0 para todas as demais classes. A expressão matemática utilizada, multiplicou então por dez os valores binários referentes a 1985 e os somou com os valores binários referentes a 2024 [(Mapa_Binário_1985 * 10) + Mapa_Binário_2024]. Esta operação resultou em um mapa de transição com quatro valores possíveis:

3.2.2.1. 11: Pixel era urbano em 1985 e permaneceu urbano em 2024.

3.2.2.2. 1: Pixel não era urbano em 1985 e tornou-se urbano em 2024.

3.2.2.3. 10: Pixel era urbano em 1985 e deixou de ser em 2024.

3.2.2.4. 0: Pixel não era urbano em 1985 e assim permaneceu em 2024.

Este mapa foi então reclassificado em três classes finais de interesse: 1 para Área Urbana Estável (antigo valor 11), 2 para Urbanização/Expansão Urbana (antigo valor 1) e 3 para Outras Classes (antigos valores 0 e 10). Adicionalmente, foi aplicada uma máscara para remover todos os pixels classificados como "Não Observado" em qualquer um dos anos, garantindo a integridade da análise.

3.2.3. Fase 3: Com o mapa de transição finalizado, foram automaticamente calculadas as áreas absolutas (em km²) e relativas (em % do estado) para cada uma das três classes. As estatísticas foram exportadas em formato de planilha (.CSV) e o mapa de transição foi exportado em formato raster (.GeoTIFF).

3.2.4. Fase 4: Ao rodar o *script* principal, também foi rodado o *script* de auditoria (camada 3), o qual produziu 500 pontos de amostragem para cada classe a fim de se obter a Verdade de Referência. O *script* de auditoria foi desenhado pelo pesquisador e a IA apenas codificou sua implementação, buscando garantir a independência do caminho metodológico para produção dos resultados em relação ao *script* principal. A lógica condicional adotada aplicou as regras abaixo:

3.2.4.1. Se era urbano em 1985 e continuou urbano em 2024 então o valor esperado será 1.

3.2.4.2. Se não era urbano em 1985 e se tornou urbano em 2024 então o valor esperado será 2.

3.2.4.3. Todos os outros casos o valor será 3.

Com todos os valores da comparação, foi gerada uma matriz de confusão, contrapondo a classificação do mapa final de transição com a Verdade de Referência, permitindo o cálculo de índices de acurácia e o coeficiente Kappa. É importante distinguir a ordem de execução computacional da ordem de validação metodológica. Embora a automação permita que o *script* de auditoria da Camada 3 seja executado simultaneamente à geração do mapa de transição — uma abordagem que otimiza o tempo de processamento —, a análise do pesquisador seguiu estritamente a sequência de defesa em profundidade. Assim, os resultados quantitativos da auditoria (matriz de confusão e índices de acurácia) só foram analisados para validar o mapa após a conclusão satisfatória das etapas anteriores: a depuração do código (Camada 1) e a inspeção visual qualitativa (Camada 2). Essa abordagem combina a eficiência computacional com o rigor da validação sequencial.

3.2.5. Fase 5: aplicou-se então a etapa dois do protocolo de segurança, com uma inspeção visual de zonas do mapa gerado em contraposição com o conhecimento prévio das áreas, atestando sua qualidade de forma também qualitativa.

3.2.6. Fase 6: A tabela de confusão e os índices de acurácia e Kappa foram analisados e aprovados, conforme previsto pela etapa três do protocolo de proteção.

3.2.7. Fase 7: implementou-se a etapa quatro do protocolo de segurança, solicitando à IA que produzisse a explicação detalhada de cada parte e funções dos códigos, a fim de permitir a auditoria “Caixa Aberta Suave” das lógicas utilizadas.

3.2.8. Fase 8: O arquivo GeoTIFF resultante da Fase 3 foi importado para o ambiente QGIS. Nesta etapa, foram aplicadas simbologia e paleta de cores para representar as classes, e outras camadas de informação (como a malha rodoviária) foram adicionadas para enriquecer a análise visual. Por fim, realizou-

se a composição do *layout* do mapa final, incluindo elementos cartográficos essenciais como legenda, escala, grade de coordenadas e título.

3.3. Outros usos da Inteligência Artificial

Em consonância com as boas práticas de integridade em pesquisa (Sampaio; Sabbatini; Limongi, 2024) e para garantir a transparência metodológica, informa-se que a ferramenta de inteligência artificial generativa Gemini 2.5 Pro foi utilizada como recurso auxiliar na preparação e elaboração deste manuscrito. As aplicações específicas incluíram:

3.3.1. Redação e Revisão: Auxílio na revisão ortográfica e no aprimoramento do estilo da escrita científica, visando maior clareza e correção gramatical.

3.3.2. Criação de Elementos Visuais: Geração da ilustração das camadas de segurança e do diagrama do protocolo de segurança (figuras 1 e 2). Ambos os elementos foram criados a partir de um processo iterativo de *prompts* fornecidos pelo pesquisador, que orientou e refinou os resultados até a versão final.

3.3.3. Tradução: Assistência na tradução para outros idiomas.

3.3.4. Uso exploratório: Realização de diálogos iniciais para explorar conceitos e estruturar ideias, cujos resultados serviram como ponto de partida para um aprofundamento subsequente em bibliografia acadêmica consolidada.

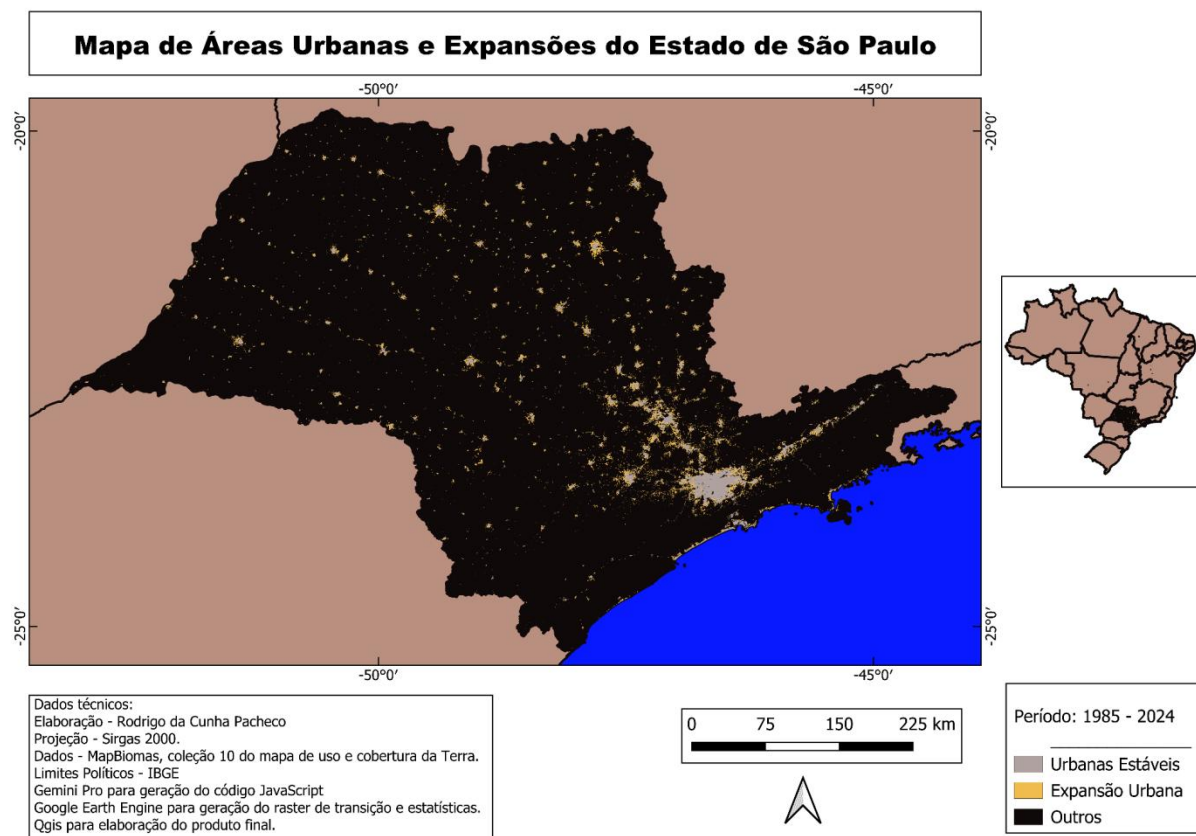
Ressalta-se que, em todas as etapas, os conteúdos e elementos gerados pela ferramenta foram tratados como dados brutos, submetidos à análise crítica, validação e edição pelo autor. O pesquisador detém a responsabilidade integral pelo conteúdo final, pelas asserções e pelos dados apresentados neste trabalho.

4. Resultado e discussão

A Figura 3 mostra o mapa de transição final obtido. Sua análise permite ao especialista um primeiro mecanismo de validação de todo o processo, afinal, caso o *script* de mapeamento e sua lógica possuíssem erros grosseiros, seria possível constatar os efeitos no produto final. No presente caso, é fácil perceber que a área central da grande mancha urbana da Região Metropolitana de São Paulo deveria ser classificada como área

urbana estável (sem alteração no período), algo que ocorre no mapa produzido, como esperado. A área do Parque Estadual Serra do Mar, criado para proteção ambiental em 1977 (Fundação Florestal, 2025), por sua vez, deveria aparecer na classe “Outras”, pois era natural e permanece natural, como novamente bem indicado no mapa. Por fim, faz sentido pensar que nas bordas de manchas urbanas estáveis, em cidades que passaram por crescimento populacional, como São Paulo, deveria haver áreas de expansão do tecido urbano, o que também se percebe no mapa. Essa análise qualitativa fornece boas indicações de sucesso no processo de mapeamento.

Figura 3. Mapa de transição das áreas urbanas do Estado de São Paulo (1985-2024).



A Tabela 1 mostra o resultado da matriz de confusão. A diagonal apresenta a quantidade de amostras classificadas corretamente. Todos os pontos de amostras do mapa de transição coincidiram com a Verdade de Referência (esperada). Isso resultou em um índice geral de acurácia de 1 e Kappa de 1. É outra evidência contundente do sucesso do mapeamento sob o protocolo aplicado.

Tabela 1. Matriz de confusão do mapa de transição

Classes	1	2	3
1	500	0	0
2	0	500	0
3	0	0	500

A Tabela 2, por sua vez, mostra os dados obtidos a respeito da extensão territorial de cada classe.

Tabela 2. Classes de transição do uso e cobertura da Terra do Estado de São Paulo (1985-2024).

Classes de Uso	Área em Ha	Área em Km²	Valores em %
Área Urbana Estável	389.585,55	3.895,9	1,57
Expansão urbana	469.025,49	4.690,3	1,89
Outras	23.963.518,50	239.635,2	96,54
Total	24.822.129,53	248.221,3	100,00

Chama a atenção o fato de a área total em km², que foi calculada a partir dos *pixels* do raster de transição do uso da terra, ter alcançado 99,9% de precisão em relação ao tamanho do estado apresentado pelo IBGE (2025b), de 248.219,485 km², o que reforça mais uma vez a robustez do código gerado.

A classe “Outras”, de longe, corresponde a maior categoria em termos de área (96,54%), o que é compreensível pelos métodos de reclassificação aplicados, focados no mapeamento da expansão urbana no período. Em um Estado relativamente grande como São Paulo, seria de esperar que as áreas rurais e de preservação ambiental ocupassem a maior parte do estado. Não obstante, houve aumento das áreas urbanas para o período de estudo, estimado em 120,39%, fazendo-as corresponderem, em 2024, à algo em torno de 3,46% do território paulista. O valor total de 8.586,2 km² de áreas urbanas totais para 2024 é levemente inferior à estimativa do IBGE para o ano de 2019 (8.614,62 km²). Certamente não houve redução dessas áreas no período, devendo a diferença ser atribuída à distinção de técnicas de mapeamento entre o IBGE e o MapBiomias (por exemplo, uso de algoritmos de

classificação heterogêneos). Por outro lado, a proximidade dos valores indica também que eventuais erros no *script* de mapeamento, caso existissem, não seriam tão grandes.

A Tabela 3 mostra os dados referentes aos habitantes do Estado de São Paulo, produzidos pelo IBGE e compilados pelo IPEA (2025), exceto para o ano de 2025, o qual corresponde a uma projeção direta do IBGE (2025b). Já a Equação 1 mostra um caminho para estimar e obter a população referente aos anos de 1985 e 2024 (anos iniciais e finais do presente estudo) por meio de uma interpolação linear para os referidos anos de interesse, assumindo um crescimento linear da população. Assim, as estimativas da população para os anos de 1985 e 2024 são respectivamente: 28.017.915 e 45.524.947 habitantes. Portanto, o crescimento estimado da população para o período foi de 62,49%.

Tabela 3. Evolução da população do Estado de São Paulo (IPEA e IBGE).

	1980	1991	1996	2000	2007	2010	2022	2025
População	25.042.074	31.588.925	34.119.110	37.035.456	39.827.570	41.262.199	44.411.238	46.081.801

Equação 1. Fórmula para estimação da população por meio de interpolação linear.

$$Y = \left[\frac{y_2 - y_1}{x_2 - x_1} \times (x - x_1) \right] + y_1$$

Onde:

Y = População estimada que queremos encontrar.

x = Ano para o qual queremos estimar a população (1985 ou 2024).

x1 e x2 = Os anos conhecidos que "cercam" o ano que queremos estimar.

y1 e y2 = As populações conhecidas correspondentes aos anos x1 e x2.

Aplicando-se a fórmula da densidade demográfica (população absoluta / área), obtém-se, respectivamente, para os anos de 1985 e 2024, a densidade demográfica estimada de 7.192 e 5.302 habitantes por quilômetro quadrado, ou queda de 26,28%. Isso ocorreu porque, embora a população do estado tenha crescido cerca de 62,49% no período, sua área urbana cresceu aproximadamente 120,39%. Assim, o estado teve um crescimento urbano de menor densidade demográfica, ou seja, de maior espraiamento do tecido urbano e, conseqüentemente, de maior exigência de construções de infraestrutura para expansão do tecido urbano. Nadalin e Iglioni indicaram a ocorrência desse processo, entre 1960 e 2000,

na Região Metropolitana de São Paulo, relacionando-o à ocupação dos subúrbios pelas famílias de baixa renda. Eles definiram o processo como “a área urbanizada cada vez menos concentrada, menos densa e com mais vazios urbanos” (Nadalin; Iglioni, 2015, p. 109). Neste estudo, porém, diferencia-se o processo de espraiamento (o crescimento) da perda de densidade populacional, a qual será denominada como processo de rarefação urbana, uma consequência do espraiamento.

No mapa da Figura 4, foi adicionada a malha viária do Estado de São Paulo, permitindo a percepção da articulação física dos caminhos de fluxos da rede urbana estadual. A Figura 5, por sua vez, foca na expansão das áreas urbanas na área aproximada da Macrometrópole Paulista, no setor mais oriental do estado, tornando mais fácil perceber o fenômeno de expansão das bordas das áreas urbanas ao redor dos tecidos mais estabilizados.

Figura 4. Mapa de transição das áreas urbanas do Estado de São Paulo (1985-2024) com a malha rodoviária.

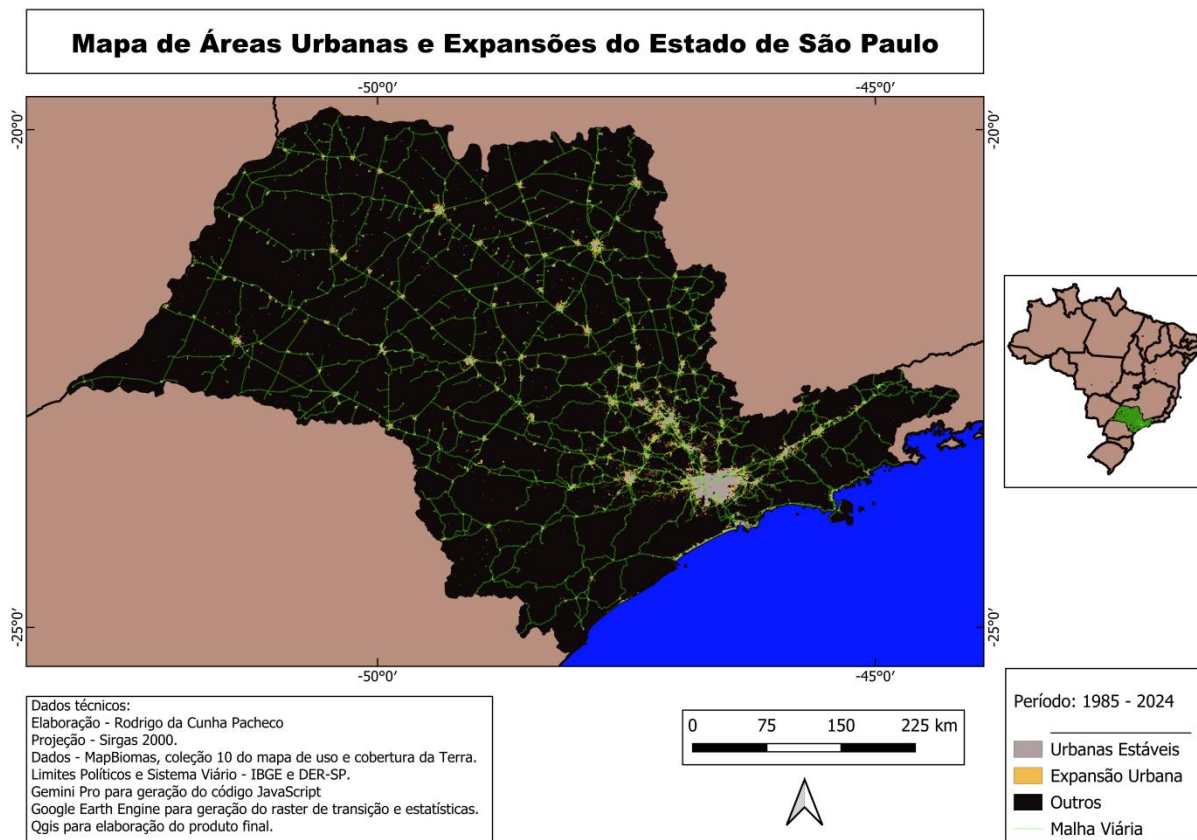
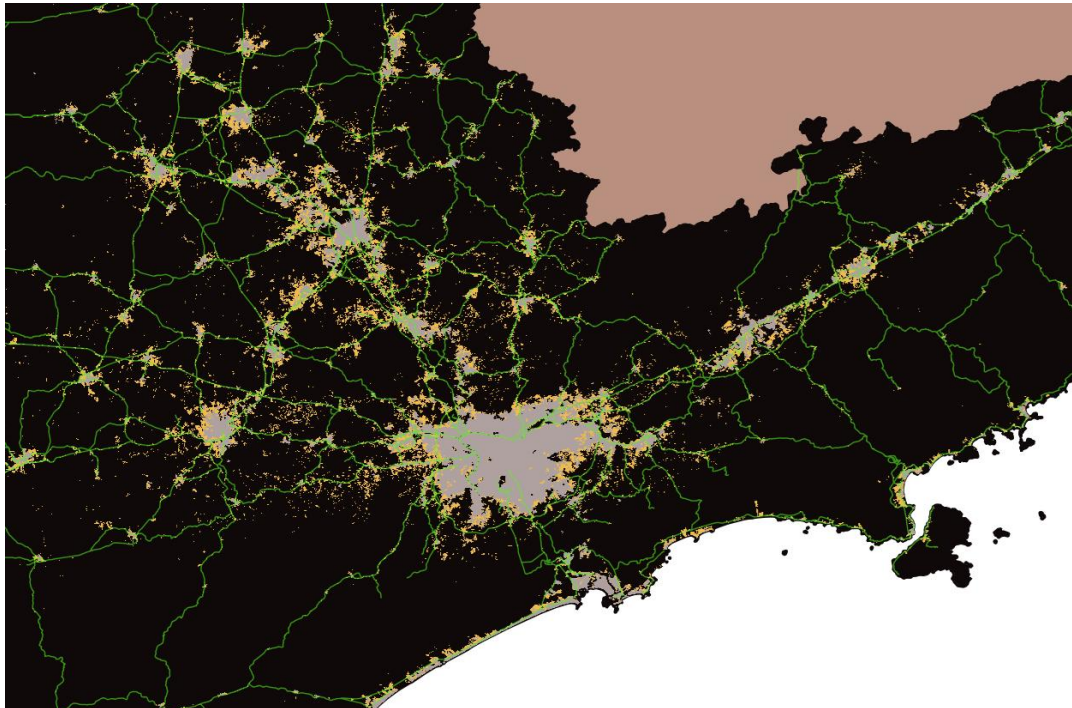


Figura 5. Captura de tela da transição das áreas urbanas do Estado de São Paulo (1985-2024), com a malha rodoviária e focada na região da Macrometrópole Paulista.



Por fim, a análise não poderia terminar sem antes abordar a questão da confiabilidade dos mapas de uso da terra de origem, os quais permitiram a elaboração do mapa de transição focado na expansão urbana. Afinal, o protocolo aqui proposto parte da pressuposição da confiabilidade dos dados de origem. No caso da coleção de cobertura da terra utilizada, esta foi produzida por uma equipe de especialistas: "O MapBiomias é uma rede global e multi-institucional, formada por universidades, ONGs e empresas de tecnologia, que monitora as transformações na cobertura e no uso da terra nos territórios e seus impactos" (MapBiomias, 2025b). Embora não tenha sido encontrada a informação de acurácia para o mapeamento da coleção 10 utilizada (provavelmente por ainda não ter sido disponibilizada), a coleção 9 apresentou um índice de 86,53% para a região do bioma da Mata Atlântica no nível 2 de mapeamento e de 91,67% no nível 1.

5. Conclusão

Esta pesquisa propôs um protocolo que permite o uso do GEE para a produção de mapeamentos de transição do uso da terra por parte de pesquisadores com conhecimentos limitados de programação, recorrendo à IA generativa, sem abrir mão da integridade científica. A aplicação do protocolo no estudo da transição urbana do Estado de São Paulo

entre os anos de 1985 e 2024, utilizando mapeamentos em raster produzidos pelo MapBiomas, foi bem-sucedida, sugerindo a sua eficácia.

A abordagem de defesa em camadas, notadamente a validação quantitativa por meio de um *script* de auditoria independente, mostrou-se fundamental para assegurar a confiabilidade do código gerado. A importância desta constatação decorre da possibilidade de promoção da acessibilidade ao GEE, por parte dos pesquisadores, com o uso responsável da IA generativa.

Os dados obtidos revelaram um crescimento considerável do tecido urbano estadual para o período de análise (120,39%), bem como uma diminuição da densidade demográfica, indicando um caráter desse processo que será conceituado como rarefação urbana, com a consequência de requerer do poder público investimentos de expansão da infraestrutura urbana em detrimento de um melhor aproveitamento da já existente.

Reconhece-se, contudo, que o protocolo foi validado para uma tarefa específica de mapeamento de transição. Pesquisas futuras poderiam, portanto, explorar sua aplicabilidade em outras análises geoespaciais no GEE, bem como testar sua eficácia com diferentes modelos de IA generativa ou em regiões com dinâmicas de uso da terra distintas.

Referências

AKINBOYEWA, T. *et al.* GIS Copilot: towards an autonomous GIS agent for spatial analysis. **International Journal of Digital Earth**, v. 18, n. 1, p. 1-23, 2025. Disponível em: <https://doi.org/10.1080/17538947.2025.2497489>. Acesso em: 05 out. 2025.

DAM, Sumit Kumar *et al.* A Complete Survey on LLM-based AI Chatbots. **arXiv**. [Pré-impressão], p. 01-23, 2024. Disponível em: <https://doi.org/10.48550/arXiv.2406.16937>. Acesso em: 05 out. 2025.

FITZ, P. R. **Geoprocessamento sem complicação**. São Paulo: Oficina de Textos, 2008.

FUNDAÇÃO FLORESTAL. **Parque Serra do Mar: quatro décadas de preservação**, 2017. Disponível em: <http://fflorestal.sp.gov.br/2017/08/parque-serra-do-mar-quatro-decadas-de-preservacao/>. Acesso em: 5 out. 2025.

GEMINI TEAM, Google. Gemini 2.5: Pushing the Frontier with Advanced Reasoning, Multimodality, Long Context, and Next Generation Agentic Capabilities. **arXiv**. [Pré-impressão], p.1-72, 2025. Disponível em: <https://doi.org/10.48550/arXiv.2507.06261>. Acesso em: 05 out. 2025.

GORELICK, N. *et al.* Google Earth Engine: Planetary-scale geospatial analysis for everyone. **Remote Sensing of Environment**, v. 202, p. 18-27, 2017. Disponível em: <https://doi.org/10.1016/j.rse.2017.06.031>. Acesso em: 05 out. 2025.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Malhas territoriais**, 2025a. Disponível em: <https://www.ibge.gov.br/geociencias/organizacao-do-territorio/malhas-territoriais/15774-malhas.html>. Acesso em: 05 out 2025.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Cidades**, 2025b. Disponível em: <https://cidades.ibge.gov.br/brasil/sp/panorama>. Acesso em: 05 out 2025.

INSTITUTO DE PESQUISA ECONÔMICA APLICADA (IPEA). **Ipeadata**, 2025. Disponível em: <https://ipeadata.gov.br/Default.aspx>. Acesso em: 05 out. 2025.

KRAKOWSKI, S. Human-AI Agency in the Age of Generative AI. **Information and Organization**. 2025, vol. 35, no. 1, p. 01-25, 2025. Disponível em: <https://doi.org/10.1016/j.infoandorg.2025.100560>. Acesso em: 05 out. 2025.

LI, Z. *et al.* GIScience in the era of Artificial Intelligence: a research agenda towards Autonomous GIS. **Annals of GIS**, p. 01-36, 2025. Disponível em: <https://doi.org/10.1080/19475683.2025.2552161>. Acesso em: 05 out. 2025.

LUDERMIR, T. B. Inteligência Artificial e Aprendizado de Máquina: estado atual e tendências [online]. **Estudos Avançados**, São Paulo, Brasil, v. 35, n. 101, p. 85–94, 2021. Disponível em: <https://revistas.usp.br/eav/article/view/185035>. Acesso em: 5 out. 2025.

MAPBIOMAS. **Coleção 10 da Série Anual de Mapas de Cobertura e Uso da Terra do Brasil**, 2025a. Disponível em: <https://brasil.mapbiomas.org/>. Acesso em: 05 out. 2025.

MAPBIOMAS. **O projeto**. 2025b. Disponível em: <https://brasil.mapbiomas.org/o-projeto/>. Acesso em: 05 out. 2025.

MORGATO, P. Geografia 2.0: para lá da tecnologia. *In*: CONGRESSO DA GEOGRAFIA PORTUGUESA, 8., 2011, Lisboa. **Actas** [...]. Lisboa: [s. n.], 2011. Disponível em: http://www.mopt.org.pt/uploads/1/8/5/5/1855409/geografia2.0_pm.pdf. Acesso em: 9 set. 2025.

NADALIN, V.; IGLIORI, D. Espreadimento urbano e periferização da pobreza na região metropolitana de São Paulo: evidências empíricas. **EURE (Santiago)**, Santiago, v. 41, n. 124, p. 91-111, 2015. Disponível em: http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0250-71612015000400005&lng=es&nrm=iso. Acesso em: 13 out. 2025.

OLIVEIRA, P.C. de. **Abordagem semi-supervisionada para detecção de módulos de software defeituosos**. 2015. 133 f. Dissertação (Mestrado em Ciência da Computação) – Centro de Informática, Universidade Federal de Pernambuco, Recife, 2015. Disponível em: <https://repositorio.ufpe.br/handle/123456789/19990>. Acesso em: 05 out. 2025.

PETERS, J. F.; PEDRYCZ, W. **Engenharia de software: teoria e prática**. Rio de Janeiro: Campus, 2001.

PFLEEGER, S. L. **Engenharia de software: teoria e prática**. 2. ed. São Paulo: Pearson Education do Brasil, 2004.

ROSA, R. Geotecnologias na geografia aplicada. **Revista do Departamento de Geografia**, São Paulo, Brasil, v. 16, p. 81-90, 2011. Disponível em: <https://doi.org/10.7154/RDG.2005.0016.0009>. Acesso em: 05 out. 2025.

RUSSELL, S. J.; NORVIG, P. **Inteligência artificial**. Rio de Janeiro: Elsevier, 2013.

SAMPAIO, R. C.; SABBATINI, M.; LIMONGI, R. **Diretrizes para o uso ético e responsável da inteligência artificial generativa**: um guia prático para pesquisadores. São Paulo: Sociedade Brasileira de Estudos Interdisciplinares da Comunicação. 2024. Disponível em: <https://www.portcom.intercom.org.br/ebooks/detalheEbook.php?id=57203>. Acesso em: 05 out. 2025.

SÃO PAULO (Estado). **Dados abertos**, 2025. Malha rodoviária estadual disponibilizada pelo DER-SP. Disponível em: <https://dadosabertos.sp.gov.br/dataset/sistema-rodoviario-estadual>. Acesso em: 5 out. 2025.

SILVA, Yuri de Melo. **Defesa em profundidade na segurança cibernética**. 2022. Trabalho de Conclusão de Curso (Graduação em Engenharia de Computação) – Faculdade de Tecnologia e Ciências Sociais Aplicadas, Centro Universitário de Brasília, Brasília, 2022. <https://repositorio.uniceub.br/jspui/handle/prefix/17287>. Acesso em: 05 out. 2025.

SOMMERVILLE, I. **Engenharia de software**. 8. ed. São Paulo: Pearson Education do Brasil, 2007.

SOUZA, J. A. C. de. **Acurácia das métricas de validação da classificação de imagens**. 2020. 39 f. Dissertação (Mestrado em Engenharia de Produção) – Escola de Engenharia, Universidade Federal de Minas Gerais, Belo Horizonte, 2020. Disponível em: <https://repositorio.ufmg.br/items/83f7df6f-dbfd-4d64-ae63-1de715ceaa32>. Acesso em: 05 out. 2025.

VASCONCELOS, S. and MARUŠIĆ, A. Integridade científica e agência humana na pesquisa entremeada por IA Generativa [online]. **SciELO em Perspectiva**, 2025. Disponível em: <https://blog.scielo.org/blog/2025/05/07/integridade-cientifica-e-agencia-humana-na-pesquisa-ia-gen/>. Acesso em: 15 out. 2025.

YANG, L. *et al.* Google Earth Engine and Artificial Intelligence (AI): A Comprehensive Review. **Remote Sensing**, v. 14, n. 14, p. 01-110, 2022. Disponível em: <https://doi.org/10.3390/rs14143253>. Acesso em: 05 out. 2025.

ZHANG, Z. *et al.* LLM Hallucinations in Practical Code Generation: Phenomena, Mechanism, and Mitigation. **Proc. ACM Softw. Eng.**, v. 2, n. ISSTA, p.1-23, 2025. Disponível em: <https://doi.org/10.1145/3728894>. Acesso em: 21 out. 2025.

Declaração de conflito de interesse

O autor declara que não há conflito de interesse.

Declaração de disponibilidade de dados da pesquisa

Todo o conjunto de dados de apoio aos resultados deste estudo foi publicado no próprio artigo e anexado ao processo de submissão na plataforma Scielo.

Este preprint foi submetido sob as seguintes condições:

- Os autores declaram que os necessários Termos de Consentimento Livre e Esclarecido de participantes ou pacientes na pesquisa foram obtidos e estão descritos no manuscrito, quando aplicável.
- Os autores declaram que a elaboração do manuscrito seguiu as normas éticas de comunicação científica.
- Os autores declaram que estão cientes que são os únicos responsáveis pelo conteúdo do preprint e que o depósito no SciELO Preprints não significa nenhum compromisso de parte do SciELO, exceto sua preservação e disseminação.
- Os autores declaram que os dados, aplicativos e outros conteúdos subjacentes ao manuscrito estão referenciados.
- O manuscrito depositado está no formato PDF.
- Os autores declaram que a pesquisa que deu origem ao manuscrito seguiu as boas práticas éticas e que as necessárias aprovações de comitês de ética de pesquisa, quando aplicável, estão descritas no manuscrito.
- Os autores declaram que uma vez que um manuscrito é postado no servidor SciELO Preprints, o mesmo só poderá ser retirado mediante pedido à Secretaria Editorial do SciELO Preprints, que afixará um aviso de retratação no seu lugar.
- Os autores concordam que o manuscrito aprovado será disponibilizado sob licença [Creative Commons CC-BY](#).
- O autor submissor declara que as contribuições de todos os autores e declaração de conflito de interesses estão incluídas de maneira explícita e em seções específicas do manuscrito.
- Os autores declaram que o manuscrito não foi depositado e/ou disponibilizado previamente em outro servidor de preprints ou publicado em um periódico.
- Caso o manuscrito esteja em processo de avaliação ou sendo preparado para publicação mas ainda não publicado por um periódico, os autores declaram que receberam autorização do periódico para realizar este depósito.
- O autor submissor declara que todos os autores do manuscrito concordam com a submissão ao SciELO Preprints.