Quando o Código Vira Commodity: O Trunfo do Matemático

Edson Cilos Vargas Júnior

Universidade Federal de Santa Catarina

2026

Contato: edson.junior@ufsc.br

Roteiro

Quando Programar Fica de “Graça”

Hackathon da Anthropic, fev/2026: 13k inscritos

Quando Programar Fica de “Graça”

Hackathon da Anthropic, fev/2026: 13k inscritos

- **1º lugar:** um advogado

Quando Programar Fica de "Graça"

Hackathon da Anthropic, fev/2026: 13k inscritos

- **1º lugar:** um advogado
- **3º lugar:** um cardiologista que construiu uma plataforma de IA para cuidados em 7 dias

Quando Programar Fica de "Graça"

Hackathon da Anthropic, fev/2026: 13k inscritos

- **1º lugar:** um advogado
- **3º lugar:** um cardiologista que construiu uma plataforma de IA para cuidados em 7 dias
- Nenhum dos dois escreveu código

Por que Desta Vez é Diferente

- Imprensa, máquinas de escrever, LaTeX: automatizaram a **disseminação**
- IA moderna: automatiza a **criação**

A pergunta "se código está se tornando commodity, o que é escasso?"
só faz sentido agora.

Klowden & Tao, arXiv:2603.26524 (2026), §2.

Quando Programar Fica de “Graça”

- Vagas de programador (foco em execução): $-27{,}5\%$ (2023–2025)
- Vagas de design (AI Engineer): $+18{,}7\%$ staff premium (mesmo período)

Quando Programar Fica de “Graça”

- Vagas de programador (foco em execução): $-27{,}5\%$ (2023–2025)
- Vagas de design (AI Engineer): $+18{,}7\%$ staff premium (mesmo período)

Se código está se tornando commodity, o que é escasso?

Roteiro

Glossário: Termos de IA

Termo	Significado
LLM	*Large Language Model* (Claude, GPT-5.4, Gemini 3)
Alucinação	Saída do modelo confiante, mas factualmente incorreta
Modelo de fronteira	Estado da arte público (Claude Opus 4.7, GPT-5.4, Gemini 3.1 Pro)
Acesso restrito	Modelo liberado com restrições (ex.: Claude Mythos, abr/2026)
Lean 4	Assistente de demonstrações formais (linguagem dependentemente tipada)
AlphaProof / AlphaEvolve	Sistemas DeepMind para pesquisa matemática (acesso restrito)
Deep Think	Modo de raciocínio prolongado do Gemini 3

Glossário: Benchmarks

Benchmark	**O que testa**
MATH-500	Ensino médio a nível olímpico, respostas em forma livre
AIME	*American Invitational Mathematics Exam* (ensino médio)
Putnam	Competição universitária dos EUA/Canadá
IMO	Olimpíada Internacional de Matemática
FrontierMath T4	Problemas em nível de pesquisa (Epoch AI)

Roteiro

IA nas Olimpíadas de Matemática

IMO 2024: Prata

- AlphaProof + AlphaGeometry 2
- Pontuação: **28/42** (a 1 ponto do ouro)
- Método: demonstrações formais em Lean 4
- Tempo: **vários dias** por problema

IMO 2025: Ouro*

- Gemini Deep Think
- Pontuação: **35/42**, 5 de 6 problemas
- Método: **linguagem natural**
- Tempo: dentro das 4h30 da prova
- **Certificado oficialmente** pela IMO

*OpenAI: 35/42 por **auto-avaliação**, sem aval da IMO. DeepSeekMath-V2: ouro reportado de forma independente.

IA nas Olimpíadas de Matemática

IMO 2024: Prata

- AlphaProof + AlphaGeometry 2
- Pontuação: **28/42** (a 1 ponto do ouro)
- Método: demonstrações formais em Lean 4
- Tempo: **vários dias** por problema

IMO 2025: Ouro*

- Gemini Deep Think
- Pontuação: **35/42**, 5 de 6 problemas
- Método: **linguagem natural**
- Tempo: dentro das 4h30 da prova
- **Certificado oficialmente** pela IMO

*OpenAI: 35/42 por **auto-avaliação**, sem aval da IMO. DeepSeekMath-V2: ouro reportado de forma independente.

De especialistas em linguagem formal para raciocínio de propósito geral em **um ano**.

AlphaGeometry 2: Especialista em Geometria

Google DeepMind, 2024: IMO-AG-50 (problemas da IMO 2000–2024)

AlphaGeometry 2: Especialista em Geometria

Google DeepMind, 2024: IMO-AG-50 (problemas da IMO 2000–2024)

- AG2 resolveu **84%** (42/50), acima da média de medalhistas de ouro (40,9)

AlphaGeometry 2: Especialista em Geometria

Google DeepMind, 2024: IMO-AG-50 (problemas da IMO 2000–2024)

- AG2 resolveu **84%** (42/50), acima da média de medalhistas de ouro (40,9)
- LLMs de propósito geral (o1, Gemini Thinking, fev/2025): **0/50**†

AlphaGeometry 2: Especialista em Geometria

Google DeepMind, 2024: IMO-AG-50 (problemas da IMO 2000–2024)

- AG2 resolveu **84%** (42/50), acima da média de medalhistas de ouro (40,9)
- LLMs de propósito geral (o1, Gemini Thinking, fev/2025): **0/50**†
- Superado por especialistas desde então:
  Seed-Geometry **43/50** (ByteDance, jul/2025) → InternGeometry **44/50** (Shanghai AI Lab, dez/2025)

†Sem reavaliação pública de GPT-5/5.4, Claude 4.x ou Gemini 3 no IMO-AG-50. O zero é dos modelos de fevereiro de 2025.

Matemática Olímpica: Quase Resolvida

Modelo	MATH-500	AIME 2025
Claude Opus 4.6	n/a	99,8%
GPT-5 / GPT-5.2	99,4%	100%*
DeepSeek-R1-0528	97,3%	93,1%
Gemini 3.1 Pro	95,1%	100%*

*Com execução de código.

MATH-500 no teto. AIME 2025 saturado. Mais a **IMO 2025: ouro**.

Matemática Olímpica: Quase Resolvida

Modelo	MATH-500	AIME 2025
Claude Opus 4.6	n/a	99,8%
GPT-5 / GPT-5.2	99,4%	100%*
DeepSeek-R1-0528	97,3%	93,1%
Gemini 3.1 Pro	95,1%	100%*

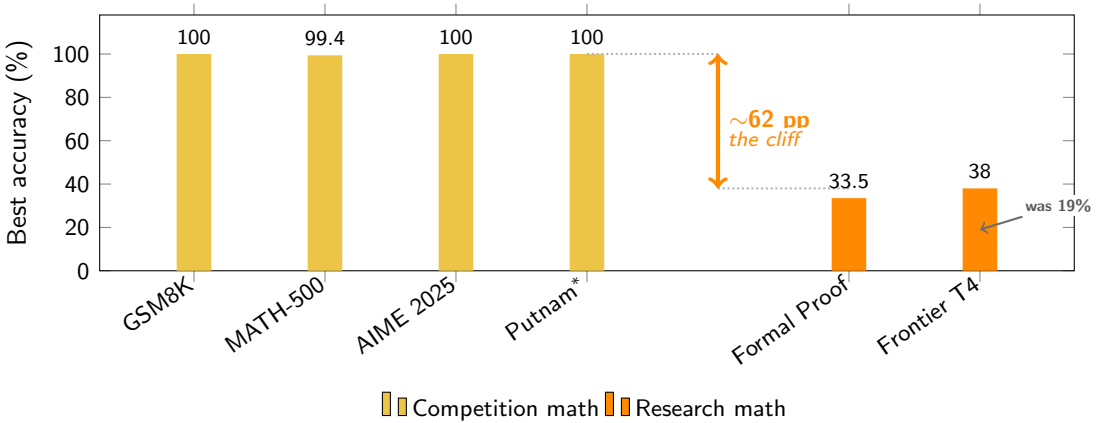
*Com execução de código.

MATH-500 no teto. AIME 2025 saturado. Mais a **IMO 2025: ouro**.

Esses benchmarks param em **nível olímpico**.

Não diferenciam mais os modelos de fronteira.

Pesquisa em Matemática: claramente em aberto



Putnam 2025: 12/12 formalmente verificado (AxiomProver, Lean 4.21)

FrontierMath T4: → 38% (GPT-5.4 Pro, mar/2026)

Onde Sua Expertise Importa

- As olimpíadas estão **quase** resolvidas.

Onde Sua Expertise Importa

- As olimpíadas estão **quase** resolvidas.

O abismo até a pesquisa **continua lá**.

Essa lacuna é onde a sua **expertise** importa.

Onde os Modelos Falham

- **Alucinação nas etapas da demonstração**
  "Uma única afirmação alucinada pode invalidar toda a cadeia de derivação"

Onde os Modelos Falham

- **Alucinação nas etapas da demonstração**
  "Uma única afirmação alucinada pode invalidar toda a cadeia de derivação"
- **Lacuna entre descrever e executar**
  Modelos descrevem algoritmos corretamente mas falham ao executá-los;
  a memória de trabalho colapsa em 20–30 ramos (linhas) paralelos **independentemente da escala**

Onde os Modelos Falham

- **Alucinação nas etapas da demonstração**
  "Uma única afirmação alucinada pode invalidar toda a cadeia de derivação"
- **Lacuna entre descrever e executar**
  Modelos descrevem algoritmos corretamente mas falham ao executá-los;
  a memória de trabalho colapsa em 20–30 ramos (linhas) paralelos **independentemente da escala**
- **Colapso de acurácia**
  Apple (2025): "colapso total de acurácia a partir de certo nível de complexidade".
  O esforço de raciocínio aumenta e depois *declina*, como se o modelo desistisse

A Lacuna entre Resposta e Demonstração

Open Proof Corpus (arXiv, jun/2025): 5.062 demonstrações olímpicas

A Lacuna entre Resposta e Demonstração

Open Proof Corpus (arXiv, jun/2025): 5.062 demonstrações olímpicas

- **6 LLMs** testados: o3, o4-mini, Gemini-2.5-Pro, Grok-3-Mini, DeepSeek-R1, Qwen3

A Lacuna entre Resposta e Demonstração

Open Proof Corpus (arXiv, jun/2025): 5.062 demonstrações olímpicas

- **6 LLMs** testados: o3, o4-mini, Gemini-2.5-Pro, Grok-3-Mini, DeepSeek-R1, Qwen3
- **$\sim$30 pontos percentuais** separam uma resposta correta de uma demonstração correta

A Lacuna entre Resposta e Demonstração

Open Proof Corpus (arXiv, jun/2025): 5.062 demonstrações olímpicas

- **6 LLMs** testados: o3, o4-mini, Gemini-2.5-Pro, Grok-3-Mini, DeepSeek-R1, Qwen3
- **$\sim$30 pontos percentuais** separam uma resposta correta de uma demonstração correta
- Agregado: **43%** das demonstrações julgadas corretas por 13 juízes humanos (2.169 / 5.062)

A Lacuna Persiste em 2026

FormalProofBench (março de 2026)

Demonstrações formais em Lean 4 (nível de pós-graduação):
melhor modelo chega a apenas **33,5%**.

Nos benchmarks que pedem só a resposta final: **saturados em ∼100%**.

Abismo inalterado.

Provas Sem Entendimento

- AlphaProof (IMO 2024): soluções **formalmente verificadas**, mas com passos redundantes ou inexplicáveis

Provas Sem Entendimento

- AlphaProof (IMO 2024): soluções **formalmente verificadas**, mas com passos redundantes ou inexplicáveis
- K&T as chamam de **"provas sem cheiro"**: passam na verificação, falham no *faro* que o matemático aplica pré-verificação

Provas Sem Entendimento

- AlphaProof (IMO 2024): soluções **formalmente verificadas**, mas com passos redundantes ou inexplicáveis
- K&T as chamam de **"provas sem cheiro"**: passam na verificação, falham no *faro* que o matemático aplica pré-verificação
- Thurston: boa matemática oferece **uma narrativa causal**, não apenas uma cadeia de implicação

Klowden & Tao (2026), §4.6 n.16 • Thurston (2006).

Cegueira de Autocrítica

A falha na auditoria da própria demonstração

LLMs julgam demonstrações de *outros* modelos com mais precisão do que as *próprias*.
Cegueira de autocrítica $\Rightarrow$ **pipelines de revisão multi-modelo** tornam-se obrigatórios.

- OPC Tabela 3: todos os modelos exceto Qwen3 piores ao avaliar a própria demonstração
- "LLMs têm dificuldade de reconhecer os próprios erros". Crítico para uso iterativo
- Implicação: **não confie em um modelo para auditar a si mesmo**

O Sinal de Trajetória

Claude Mythos Preview (abril de 2026)

Anthropic anuncia: **restrito a ∼40–52 organizações** via Project Glasswing.

- USAMO 2026 (seletiva da equipe EUA para a IMO): **97,6%**‡
- **Primeiro modelo de fronteira divulgado publicamente e retido de GA***

*GA (*General Availability*): etapa de lançamento em que o produto fica disponível ao público em geral, sem lista de espera nem acesso restrito. Retido de GA = anunciado, mas ainda não liberado.

‡Pass@avg em 10 tentativas por problema, esforço máximo + raciocínio adaptativo, orçamento ilimitado (system card Mythos §6.8).

Roteiro

A Crise do Desenvolvedor Júnior

Estudo	Dados	Achado	Período
Stanford	Folhas ADP	Devs 22–25: **−20%** do pico	Ago/2025
Harvard	62M trabalhadores	Vagas júnior em IA: **−7,7%**	2025
Northeastern	Registros de emprego	Razão jr/sr: **−16,3%**	2025
Reino Unido	Vagas tech	Entry-level: **−46%**	2024
NY Fed	Desemprego nacional	Graduados em CC: **∼6,1%**	T1/2026

- 54% dos líderes de engenharia esperam menos contratações júnior a longo prazo

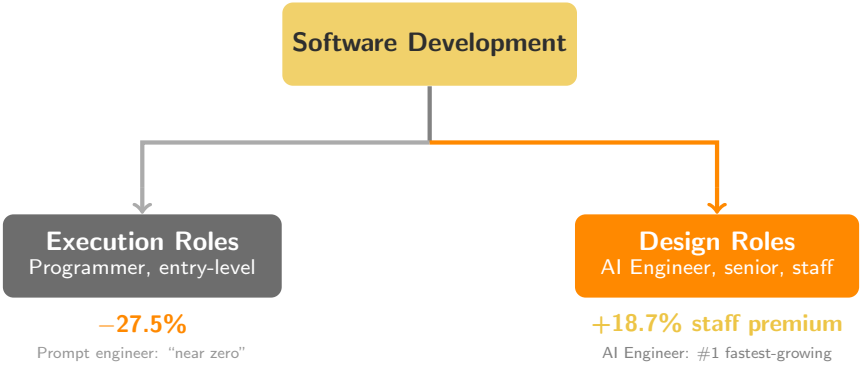
Contraevidência: Federal Reserve (março/2026, 1M+ empresas) não achou **vínculo significativo entre IA e contratações**; o mecanismo é contestado.

Empresas Reduzindo Equipes de Engenharia

Empresa	**Ação**	**Data**
Salesforce	Zero novos engenheiros contratados (AF2026)	Jan/2026
Shopify	Precisa primeiro provar que a IA não faz	Abr/2025
Microsoft	$\sim$6.000 demissões; 20–30% do código por IA	Mai/2025
Amazon	14.000 cortes corporativos; IA citada	Out/2025
Klarna	Atrito 5,5k$\rightarrow$3k; **reversão parcial**	2023–2025

Ressalva de atribuição: as empresas não publicaram comparações rigorosas de produção antes/depois. Fala de CEO $\neq$ evidência auditada.

Bifurcação, Não Extinção

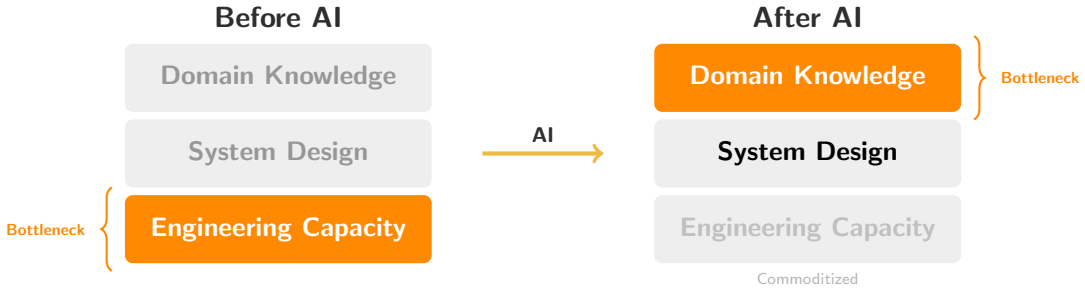


O BLS* ainda projeta **15% de crescimento** para desenvolvedores até 2034.
O agregado esconde a mudança de composição.

*BLS: *U.S. Bureau of Labor Statistics*, agência de estatísticas do trabalho dos EUA.

O Verdadeiro Gargalo

"O problema central agora é saber que coisa você, de fato, quer construir."
parafraseado de Boris Cherny, Head of Claude Code
(Lenny's Podcast, fevereiro de 2026)



Onde o Raciocínio Matemático Vale Ouro

Área	**Por que a matemática importa**
Segurança de IA (Alinhamento)	Reward hacking, dinâmica de treino
Verificação Formal	Demonstrações em Lean, verificação de IA
Teoria de ML	Interpretabilidade, generalização
Quant (mercado financeiro)	Modelos estocásticos + ML
Pesquisa Operacional	**Formular o problema** (LP/MIP)

- Áreas em que **rigor e demonstração não podem ser automatizados**
- A IA amplifica o matemático; não substitui o julgamento

O Faro Matemático

Matemáticos experientes detectam argumentos ruins **antes** de verificá-los linha a linha. Uma habilidade que LLMs atuais não têm por construção.

É o que vocês já têm.

As "Dez Pistas" de Aaronson $\rightarrow$ seu instinto pré-verificação.

Aaronson (2008) • Klowden & Tao (2026), §4.2.

Roteiro

Tao × DeepMind (nov/2025)

Terence Tao (UCLA, Medalha Fields 2006) colabora com a Google DeepMind

- Tao *seleciona e avalia* os problemas; DeepMind aplica AlphaEvolve + Gemini Deep Think

Tao × DeepMind (nov/2025)

Terence Tao (UCLA, Medalha Fields 2006) colabora com a Google DeepMind

- Tao *seleciona e avalia* os problemas; DeepMind aplica AlphaEvolve + Gemini Deep Think
- **67 problemas em aberto** testados em combinatória, análise e teoria dos números

Tao × DeepMind (nov/2025)

Terence Tao (UCLA, Medalha Fields 2006) colabora com a Google DeepMind

- Tao *seleciona e avalia* os problemas; DeepMind aplica AlphaEvolve + Gemini Deep Think
- **67 problemas em aberto** testados em combinatória, análise e teoria dos números
- **$\sim$20 resultados inéditos**, superando os melhores limites conhecidos na literatura

Tao × DeepMind (nov/2025)

Terence Tao (UCLA, Medalha Fields 2006) colabora com a Google DeepMind

- Tao *seleciona e avalia* os problemas; DeepMind aplica AlphaEvolve + Gemini Deep Think
- **67 problemas em aberto** testados em combinatória, análise e teoria dos números
- **$\sim$20 resultados inéditos**, superando os melhores limites conhecidos na literatura
- Parte das demonstrações passou por **verificação formal em Lean 4**

Fonte: Tao, *Mathematical exploration and discovery at scale* (blog pessoal, 2025-11-05).

Ernest Ryu (Seoul National University; antes UCLA)

- Pesquisador em otimização convexa: **operator splitting** e métodos de primeira ordem

Ernest Ryu (Seoul National University; antes UCLA)

- Pesquisador em otimização convexa: **operator splitting** e métodos de primeira ordem
- Problema em aberto há $\sim$**40 anos** sobre taxas de convergência

Ryu (SNU, ago/2025): Otimização Convexa

Ernest Ryu (Seoul National University; antes UCLA)

- Pesquisador em otimização convexa: **operator splitting** e métodos de primeira ordem
- Problema em aberto há $\sim$**40 anos** sobre taxas de convergência
- Usou GPT-5 como assistente $\rightarrow$ **melhoria de $\sim$50%** no melhor limite conhecido

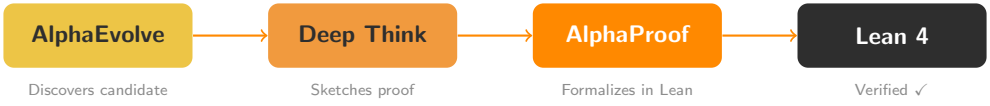
Ryu (SNU, ago/2025): Otimização Convexa

Ernest Ryu (Seoul National University; antes UCLA)

- Pesquisador em otimização convexa: **operator splitting** e métodos de primeira ordem
- Problema em aberto há $\sim$**40 anos** sobre taxas de convergência
- Usou GPT-5 como assistente $\rightarrow$ **melhoria de $\sim$50%** no melhor limite conhecido
- **Verificado e publicado**: resultado rigoroso, não apenas uma heurística

Fonte: OpenAI, *GPT-5 mathematical discovery* (anúncio oficial, agosto de 2025); nota técnica publicada por E. Ryu (SNU).

O pipeline (Tao × DeepMind) que produziu ∼*20 resultados inéditos*:



Fazer Software Sem Escrever Código

Software 3.0 de Andrej Karpathy (ex-OpenAI / ex-Tesla AI; jun/2025)

- **1.0:** código que humanos escrevem

Fazer Software Sem Escrever Código

Software 3.0 de Andrej Karpathy (ex-OpenAI / ex-Tesla AI; jun/2025)

- **1.0:** código que humanos escrevem
- **2.0:** modelos que humanos treinam

Fazer Software Sem Escrever Código

Software 3.0 de Andrej Karpathy (ex-OpenAI / ex-Tesla AI; jun/2025)

- **1.0:** código que humanos escrevem
- **2.0:** modelos que humanos treinam
- **3.0:** programas que humanos **descrevem em linguagem natural**

Não-Desenvolvedores em Produção

- Claude Code, Cursor: descreva o que você quer, receba software funcional

Não-Desenvolvedores em Produção

- Claude Code, Cursor: descreva o que você quer, receba software funcional
- Não-desenvolvedor na Arkance: app web com IA sozinho em **48 horas** (Cursor, AU 2025)

Não-Desenvolvedores em Produção

- Claude Code, Cursor: descreva o que você quer, receba software funcional
- Não-desenvolvedor na Arkance: app web com IA sozinho em **48 horas** (Cursor, AU 2025)
- Não-desenvolvedores já publicam apps em produção com bancos, cobrança e APIs

A sua expertise é a linguagem que importa.

Como Usar IA Hoje

Para pesquisa na fronteira, no estágio atual da IA:

- **Red team, não blue team.**
  Deixe a IA verificar seu trabalho, não criar o que você não consegue checar.
- **Faro primeiro.**
  Ceticismo pré-verificação é o seu diferencial.

Klowden & Tao (2026), §4.2, §6.2.

Roteiro

De “Pós-Graduando Mediano” a “Pronto para o Horário Nobre”

Isso é em **18 meses**.

O Trunfo do Matemático

Klowden & Tao: a IA é segura como **red team** (revisar, encontrar erros); mas perigosa como **blue team** (criar) sem alguém capaz de verificar.

E mesmo dentro do red team, o **faro matemático**,
que detecta o argumento ruim *antes* da checagem linha a linha,
é uma habilidade que LLMs não têm por construção.

Esse é o seu **trunfo na fronteira do conhecimento**.

Klowden & Tao (2026), §4.2, §6.2.

Você É o Protagonista

Os melhores modelos disponíveis quando vocês terminarem o doutorado **ainda não existem**.

Você É o Protagonista

Os melhores modelos disponíveis quando vocês terminarem o doutorado **ainda não existem**.

Mas ainda vão precisar de alguém capaz de distinguir
uma demonstração válida de uma alucinada.

Esse é você.

Obrigado

Obrigado. Perguntas?

edson.junior@ufsc.br