Revista ft

www.revistaft.com.br

Revista FT | ISSN 1678-0817 | v. 30, n. 163, 2026 | DOI: 10.69849/gm2sb559

Produtividade líquida no empreendedorismo individual: quanto a IA economiza depois da revisão e do retrabalho?

Ensaio teórico com proposições e agenda de pesquisa

Net productivity in individual entrepreneurship: how much time does ai save after review and rework?
A theoretical essay with propositions and a research agenda

Dani Hayoshi[1]

Resumo

A promessa mais difundida da inteligência artificial generativa (IAG) para empreendedores individuais é a economia de tempo. A evidência experimental registra ganhos expressivos em tarefas delimitadas, mas estudos de adoção em larga escala encontram economias autorrelatadas modestas, e ao menos um ensaio randomizado com profissionais experientes encontrou aumento, e não redução, do tempo de execução. Este ensaio teórico argumenta que parte dessa discrepância decorre de uma confusão entre produtividade bruta — o tempo poupado na geração de um produto — e produtividade líquida — o tempo e o valor poupados depois de descontados a especificação da tarefa, a verificação, a correção, o retrabalho, as trocas de contexto e o custo esperado dos erros que escapam à revisão. Para o empreendedor individual, essa distinção é especialmente relevante, porque a mesma pessoa gera, revisa, corrige e responde pelos erros, sem divisão de trabalho que absorva esses custos. Integrando a economia da produtividade e da alocação do tempo, a literatura sobre custos da qualidade e retrabalho, a pesquisa sobre automação e dependência excessiva, os estudos de interação humano-IA e a evidência empírica recente sobre IAG no trabalho, propomos um modelo contábil da produtividade líquida, com componentes de custo explícitos e condições de contorno, e derivamos oito proposições. Discutimos implicações para empreendedores individuais, limitações e um desenho de estudo com randomização no nível da tarefa e um cálculo de poder ilustrativo. O artigo não apresenta dados empíricos; suas proposições destinam-se a teste.

[1] Pesquisadora independente, São José do Rio Preto, SP, Brasil - E-mail para correspondência: dani@danihayoshi.com

Palavras-chave: produtividade; inteligência artificial generativa; empreendedorismo individual; retrabalho; verificação; custo da qualidade.

Abstract

The most widespread promise of generative artificial intelligence (GenAI) for solo entrepreneurs is time savings. Experimental evidence documents substantial gains on well-defined tasks, but large-scale adoption studies find modest self-reported savings, and at least one randomized trial with experienced professionals found an increase, rather than a reduction, in completion time. This theoretical essay argues that part of this discrepancy stems from conflating gross productivity — the time saved in generating an output — with net productivity — the time and value saved after accounting for task specification, verification, correction, rework, context switching, and the expected cost of errors that escape review. For solo entrepreneurs, this distinction is especially relevant because the same person generates, reviews, corrects, and answers for errors, with no division of labor to absorb these costs. Integrating the economics of productivity and time allocation, the literature on quality costs and rework, research on automation and overreliance, human–AI interaction studies, and recent empirical evidence on GenAI at work, we propose an accounting model of net productivity, with explicit cost components and boundary conditions, and derive eight propositions. We discuss implications for solo entrepreneurs, limitations, and a study design with task-level randomization and an illustrative power calculation. The article presents no empirical data; its propositions are intended for testing.

Keywords: productivity; generative artificial intelligence; solo entrepreneurship; rework; verification; cost of quality.

1 Introdução

"Quanto tempo a IA me economiza?" é, provavelmente, a pergunta mais frequente de quem trabalha sozinho e começa a usar ferramentas de inteligência artificial generativa. A resposta mais comum — a de que a IA escreve, resume, planeja e programa em segundos o que levaria horas — mede apenas uma parte do trabalho. Um texto gerado em segundos ainda precisa ser lido, conferido e ajustado; uma planilha precisa ter suas fórmulas checadas; um trecho de código precisa ser testado; uma resposta a um cliente precisa ser revisada quanto a fatos, tom e compromissos assumidos. Quando a revisão falha, o erro reaparece mais tarde, como retrabalho, reclamação ou perda de confiança.

A literatura empírica oferece um quadro aparentemente contraditório. Em experimentos com tarefas delimitadas, a IAG reduziu o tempo e aumentou a qualidade de textos profissionais (Noy & Zhang, 2023), aumentou o número de atendimentos resolvidos por hora em centrais de suporte (Brynjolfsson et al., 2025), acelerou a implementação de uma tarefa de programação especificada em detalhe (Peng et al., 2023) e aumentou o número de tarefas concluídas por desenvolvedores em três experimentos de campo corporativos (Cui et al., 2026). Em contraste, o único ensaio randomizado com donos de pequenos negócios que localizamos não encontrou efeito médio da IA sobre receitas e lucros (Otis et al., 2026), levantamentos representativos encontraram economias autorrelatadas da ordem de poucos pontos percentuais das horas de trabalho (Bick et al., 2026; Humlum & Vestergaard, 2025), e um ensaio randomizado com desenvolvedores experientes trabalhando em seus próprios projetos encontrou que permitir o uso de IA aumentou o tempo de conclusão das tarefas em 19%, embora os participantes acreditassem, mesmo depois do estudo, ter ficado cerca de 20% mais rápidos (J. Becker et al., 2025).

Essas diferenças têm muitas explicações possíveis: tipos de tarefa, nível de experiência, maturidade das ferramentas, desenho dos estudos, medidas de resultado. Este ensaio examina uma delas, que consideramos subteorizada: a distinção entre **produtividade bruta** e **produtividade líquida**. Produtividade bruta é o ganho observado na etapa de geração. Produtividade líquida é o ganho que sobra depois de contabilizados todos os custos que o uso da IA cria ou desloca — especificar a tarefa, esperar e interagir, verificar, corrigir, refazer, alternar entre tarefas e arcar com os erros não detectados. A pergunta central do artigo é: **quanto a IA economiza para o empreendedor individual depois da revisão e do retrabalho, e de que depende essa economia?**

O foco no empreendedor individual não é acidental. Em organizações, os custos de verificação e retrabalho podem ser distribuídos entre funções — revisores, controle de qualidade, suporte —, e os erros podem ser absorvidos por processos e equipes. No empreendedorismo individual, a mesma pessoa gera o conteúdo, revisa, corrige, entrega ao cliente e responde pelos erros. O tempo gasto em verificação é retirado diretamente de outras atividades do negócio, e seu custo de oportunidade é o da própria empreendedora (G. S. Becker, 1965). Além disso, muitos pequenos negócios operam com objetivos que não se resumem à maximização do crescimento (Hurst & Pugsley, 2011), o que torna o tempo liberado — e não apenas o volume produzido — uma medida relevante de resultado.

A motivação prática deste ensaio vem da experiência da autora com a formação de líderes e empreendedores por meio do método AI Thinking, organizado em três blocos (Questionar, Redesenhar e Reconstruir). Na prática de formação, a pergunta sobre a economia real de tempo aparece com frequência e raramente tem resposta precisa. O artigo não avalia o método nem pressupõe sua eficácia; trata a questão exclusivamente com base na literatura e formula proposições refutáveis.

O artigo oferece três contribuições. Primeiro, propõe um modelo contábil da produtividade líquida que torna explícitos os componentes de custo frequentemente omitidos nas avaliações de uso da IA, incluindo o custo esperado de erros residuais. Segundo, especifica condições sob as quais a produtividade líquida tende a ser alta, baixa ou negativa — em particular, a assimetria entre custo de gerar e custo de verificar, a expertise do usuário e a fragmentação do trabalho. Terceiro, traduz o modelo em oito proposições e em um desenho de estudo que mede separadamente cada componente de custo.

O texto está organizado da seguinte forma. A Seção 2 revisa a literatura. A Seção 3 apresenta o modelo conceitual. A Seção 4 desenvolve as proposições. A Seção 5 discute implicações para empreendedores individuais. A Seção 6 trata das limitações. A Seção 7 propõe uma agenda de pesquisa e um desenho de estudo. A Seção 8 conclui.

2 Referencial teórico

2.1 Produtividade, alocação do tempo e o paradoxo da produtividade

Na economia, produtividade é, em sua forma mais simples, a razão entre produto e insumos utilizados para obtê-lo (Syverson, 2011). Essa definição, aparentemente trivial, já contém duas advertências relevantes para este ensaio. A primeira é que o produto deve ser comparável: medir apenas a quantidade, sem ajustar pela qualidade, pode produzir ganhos ilusórios. A segunda é que todos os insumos devem ser contados: se uma tecnologia reduz o tempo de uma etapa, mas cria trabalho em outra, a produtividade só aumenta se o saldo for positivo.

G. S. Becker (1965) formalizou o tempo como recurso escasso, com custo de oportunidade, alocado entre atividades de produção e consumo. Para o empreendedor individual, essa abordagem implica que uma hora gasta revisando um texto gerado por IA não é neutra: ela compete com horas de atendimento, venda, produção ou descanso. O valor da economia de tempo depende, portanto, do uso alternativo desse tempo.

A história da adoção de tecnologias de informação sugere cautela. Brynjolfsson (1993) analisou o chamado paradoxo da produtividade da tecnologia de informação — a dificuldade de encontrar, nos dados agregados, ganhos proporcionais aos investimentos — e organizou as explicações em quatro grupos: erros de mensuração, defasagens temporais, redistribuição de ganhos e má gestão da tecnologia. Brynjolfsson et al. (2021) propuseram a "curva J" da produtividade: tecnologias de propósito geral exigem investimentos complementares intangíveis (novos processos, treinamento, reorganização do trabalho) que, no início, reduzem a produtividade medida e só depois a aumentam. Acemoglu (2025), em análise macroeconômica, argumentou que os ganhos agregados da IA na década seguinte tendem a ser modestos, dado o conjunto de tarefas efetivamente afetadas e os ganhos de custo observados nelas.

Essas contribuições operam no nível da economia ou da firma, mas sua lógica se aplica ao nível individual: o empreendedor que adota IA também investe em aprendizagem, também reorganiza seu trabalho e também pode medir mal os próprios ganhos.

2.2 Custos da qualidade e retrabalho

A gestão da qualidade desenvolveu, ao longo do século XX, modelos de custo da qualidade que distinguem custos de prevenção, de avaliação (inspeção e verificação), de falhas internas (defeitos encontrados antes da entrega) e de falhas externas (defeitos encontrados pelo cliente). Schiffauerova e Thomson (2006), em revisão da pesquisa sobre esses modelos, mostram que o modelo de prevenção, avaliação e falha é o mais difundido e que a lógica central dessa tradição é a de que investir em prevenção e avaliação reduz custos de falha, até certo ponto.

Na engenharia de software, essa lógica foi quantificada de maneira influente. Boehm e Basili (2001), em lista de dez constatações sobre redução de defeitos, afirmaram que encontrar e corrigir um problema depois da entrega costuma ser muito mais caro do que fazê-lo nas fases de requisitos e projeto, e que uma parcela substancial do esforço dos projetos de software era consumida por retrabalho evitável. Independentemente da magnitude exata em cada contexto, o princípio é robusto: o custo de um erro cresce com a distância entre o momento em que ele é introduzido e o momento em que é detectado.

Transposto para o uso da IA, o modelo de custo da qualidade sugere uma decomposição natural. A especificação cuidadosa da tarefa é custo de prevenção. A leitura e a checagem das respostas são custo de avaliação. A correção antes da entrega é custo de falha

interna. O erro que chega ao cliente — um dado incorreto em uma proposta, uma cláusula inadequada, uma promessa que não pode ser cumprida — é custo de falha externa. Uma avaliação da produtividade da IA que considere apenas a velocidade de geração ignora as quatro categorias.

2.3 As ironias da automação e a dependência excessiva

Bainbridge (1983) descreveu as "ironias da automação": ao automatizar as partes mais fáceis de um trabalho, os projetistas deixam aos humanos as partes mais difíceis — monitorar o sistema e intervir quando ele falha —, justamente as que os humanos executam pior quando a automação funciona bem na maior parte do tempo. Além disso, a falta de prática manual degrada as habilidades necessárias para intervir. Essa análise, formulada para sistemas industriais, antecipa com precisão um problema central do uso da IAG: o trabalho humano desloca-se da produção para a verificação.

A literatura sobre automação distinguiu o uso apropriado, o uso indevido (confiança excessiva), o desuso (subutilização) e o abuso (Parasuraman & Riley, 1997) e descreveu o viés de automação como a tendência a aceitar recomendações automatizadas sem verificação suficiente, produzindo erros de omissão e de comissão (Parasuraman & Manzey, 2010). Goddard et al. (2012) identificaram fatores que agravam e atenuam esse viés em sistemas de apoio à decisão clínica, entre eles a carga de trabalho e a confiança no sistema.

No caso da IAG, o problema é agravado pela natureza das respostas. Modelos de linguagem podem produzir conteúdo fluente e plausível, mas incorreto ou sem fundamento nas fontes, fenômeno conhecido como alucinação (Ji et al., 2023). Mesmo ferramentas especializadas e comercializadas como confiáveis apresentam essas falhas: em avaliação pré-registrada de ferramentas de pesquisa jurídica baseadas em IA, Magesh et al. (2025) encontraram respostas alucinadas em uma proporção que variou de cerca de 17% a 33% das consultas, conforme a ferramenta. Em programação, Pearce et al. (2022) encontraram que cerca de 40% dos programas gerados por um assistente de código em cenários relevantes para segurança continham vulnerabilidades, e Perry et al. (2023) encontraram, em estudo com usuários, que participantes com acesso a um assistente de IA escreveram código menos seguro e, ao mesmo tempo, acreditavam mais na segurança de seu código.

A dependência excessiva tem consequências cognitivas. Lee et al. (2025), em levantamento com 319 trabalhadores do conhecimento, encontraram que maior confiança na IA estava associada a menos pensamento crítico relatado e que o esforço cognitivo se

deslocava da execução para a verificação e a integração das respostas. Bastani et al. (2025) encontraram, em experimento com estudantes, que o acesso irrestrito à IA melhorou o desempenho durante a prática, mas piorou o desempenho posterior sem a ferramenta. Buçinca et al. (2021) mostraram que intervenções que obrigam o usuário a pensar antes de ver a recomendação reduzem a dependência excessiva, ao custo de uma experiência percebida como menos agradável — isto é, a verificação de qualidade tem custo, inclusive subjetivo.

2.4 Evidência empírica sobre IA generativa no trabalho

Ganhos em tarefas delimitadas. Noy e Zhang (2023) atribuíram aleatoriamente o acesso ao ChatGPT a 453 profissionais em tarefas de escrita específicas de suas ocupações e encontraram redução do tempo e aumento da qualidade avaliada. Brynjolfsson et al. (2025) encontraram, entre milhares de agentes de atendimento, aumento médio de cerca de 15% nos problemas resolvidos por hora, concentrado entre os menos experientes; para os mais experientes, os ganhos foram pequenos. Peng et al. (2023) encontraram, em experimento controlado, que desenvolvedores com acesso a um assistente de código concluíram uma tarefa de implementação bem especificada substancialmente mais rápido. Cui et al. (2026), combinando três experimentos de campo com 4.867 desenvolvedores em grandes empresas, estimaram aumento de 26,08% no número de tarefas concluídas entre os que usaram a ferramenta, com maiores ganhos entre os menos experientes. Dell'Acqua, McFowland, et al. (2026) encontraram, em experimento com 758 consultores, ganhos de produtividade e qualidade em tarefas dentro da fronteira de capacidade do modelo e piora da correção em uma tarefa fora dela. Em experimento com 791 profissionais de uma empresa de bens de consumo, Dell'Acqua, Ayoubi, et al. (2026) encontraram que indivíduos com IA alcançaram desempenho comparável ao de equipes sem IA em desafios de inovação de produto. No mesmo estudo, porém, os participantes com IA acertaram menos ao escolher a melhor entre as próprias ideias (cerca de 37%, contra cerca de 50% nas equipes sem IA): a IA elevou a qualidade do que foi gerado sem melhorar a avaliação do que foi gerado.

Economias modestas em uso real. Bick et al. (2026), com base em levantamentos representativos nos Estados Unidos, estimaram que os usuários de IAG no trabalho relataram economia média de cerca de 5,4% de suas horas de trabalho na semana anterior, o que corresponde a cerca de 1,4% das horas de todos os trabalhadores quando se incluem os não usuários. Humlum e Vestergaard (2025), vinculando levantamentos a registros administrativos na Dinamarca, relataram, na versão de maio de 2025 de seu trabalho, economia média autorrelatada de 2,8% das horas de trabalho entre usuários de chatbots e efeitos

estatisticamente nulos sobre salários e horas registradas. Dillon et al. (2025), em experimento de campo com mais de sete mil trabalhadores do conhecimento em 66 empresas, encontraram redução do tempo dedicado ao e-mail, mas não encontraram mudanças significativas no tempo de reuniões ou de redação de documentos.

Evidência com empreendedores. O único ensaio randomizado com donos de pequenos negócios que localizamos aponta na mesma direção. Otis et al. (2026) deram a 640 empreendedores quenianos acesso a um assistente de negócios baseado em GPT-4 e não encontraram efeito médio sobre receitas e lucros. A média escondia efeitos opostos: os empreendedores de baixo desempenho inicial tiveram resultados cerca de 10% piores, e os de alto desempenho, ganhos que podem ter superado 15%, embora apenas sugestivos do ponto de vista estatístico. A diferença não estava nas perguntas feitas nem nos conselhos recebidos, mas em quais conselhos cada um escolheu implementar: os de alto desempenho adaptavam as recomendações ao próprio negócio, enquanto os de baixo desempenho adotavam com mais frequência estratégias genéricas, como descontos e publicidade. O estudo mede resultados do negócio, e não tempo, mas sugere que o valor da IA para o empreendedor depende menos do que ela gera do que da capacidade de quem a usa de julgar o que foi gerado — o componente de verificação deste ensaio.

Perdas e lacunas de percepção. O estudo de J. Becker et al. (2025) merece atenção especial. Dezesseis desenvolvedores experientes, trabalhando em 246 tarefas reais de projetos que conheciam profundamente, tiveram cada tarefa aleatoriamente designada para permitir ou proibir o uso de IA. Com IA, as tarefas levaram, em média, 19% mais tempo. Antes do estudo, os participantes previam ganhar 24% de velocidade; depois, estimavam ter ganho 20%. Os próprios autores tratam o resultado como um retrato das ferramentas do início de 2025, em um contexto específico — profissionais muito experientes, projetos maduros e com altos padrões de qualidade —, e, em atualização posterior, relataram dificuldades metodológicas para medir o efeito de ferramentas mais recentes, em razão de efeitos de seleção e da dificuldade de registrar o tempo quando os desenvolvedores passaram a fazer outras tarefas enquanto agentes de IA executavam as suas (J. Becker et al., 2026). Ainda assim, o estudo documenta dois fenômenos centrais para este ensaio: a possibilidade de produtividade líquida negativa em certos contextos e a distância entre o ganho percebido e o ganho medido.

Essa distância não é exclusiva da IA. A pesquisa sobre a falácia do planejamento mostra que as pessoas subestimam sistematicamente o tempo necessário para concluir tarefas,

mesmo quando sabem que tarefas semelhantes atrasaram no passado (Buehler et al., 1994). E, em ferramentas de programação, a taxa de aceitação de sugestões — uma medida de fluidez da interação — foi o melhor preditor da produtividade autorrelatada pelos usuários (Ziegler et al., 2024), o que sugere que a experiência de "receber algo pronto" pode pesar mais na percepção do que o tempo total efetivamente gasto.

2.5 Onde o tempo vai: verificação, interação e troca de contexto

Estudos de interação humano-IA começaram a decompor o tempo de uso. Mozannar et al. (2024), ao registrar e rotular o comportamento de programadores usando um assistente de código, encontraram que a verificação de sugestões foi o estado que mais consumiu tempo nas sessões, respondendo por cerca de 22% do tempo em média. Barke et al. (2023) descreveram dois modos de uso: um modo de aceleração, em que o programador sabe o que quer e usa a ferramenta para completar rapidamente, e um modo de exploração, em que está incerto e usa a ferramenta para descobrir alternativas, com interação mais lenta e validação mais cuidadosa. Vaithilingam et al. (2022) observaram que, embora a maioria dos participantes preferisse usar o assistente, ele não melhorou significativamente o tempo de conclusão das tarefas, e que os usuários tinham dificuldade para entender, editar e depurar trechos extensos de código gerado.

Esses achados apontam para custos que não aparecem na etapa de geração. Primeiro, há o custo de **especificar** a tarefa: a interação com modelos de linguagem impõe demandas metacognitivas de explicitação de objetivos e de decomposição da tarefa (Tankelevitch et al., 2024). Segundo, há o custo de **verificar**, que depende da natureza do produto: conferir um resumo de um texto curto que se conhece é barato; conferir uma análise financeira com premissas ocultas é caro. Terceiro, há o custo de **trocar de contexto**. A pesquisa em psicologia cognitiva documenta custos de desempenho na alternância entre tarefas (Monsell, 2003), e Leroy (2009) mostrou que a atenção residual de uma tarefa anterior reduz o desempenho na tarefa seguinte. Mark et al. (2008) encontraram que pessoas interrompidas completavam tarefas mais rápido, mas com mais estresse, frustração e esforço percebido — um custo que não aparece nas medidas de tempo. Para quem trabalha sozinho e alterna entre muitos papéis, o ciclo de pedir, esperar, conferir e corrigir pode fragmentar o trabalho de maneira significativa. Por fim, a teoria da carga cognitiva sugere que tarefas de avaliação exigentes competem pelos mesmos recursos limitados de memória de trabalho usados na resolução de problemas (Sweller, 1988).

2.6 Especificidades do empreendedorismo individual

Três características distinguem o empreendedor individual dos trabalhadores estudados na maior parte da literatura revisada. A primeira é a **ausência de divisão do trabalho na verificação**. Em organizações, a revisão pode ser atribuída a outra pessoa, a um processo de controle de qualidade ou ao próprio cliente interno. No trabalho individual, a pessoa que gera o produto com a IA é a mesma que precisa julgá-lo, o que elimina a possibilidade de uma segunda leitura independente e aumenta a exposição ao viés de automação descrito por Parasuraman e Manzey (2010).

A segunda é a **amplitude de funções**. O empreendedor individual atua, frequentemente no mesmo dia, em produção, vendas, atendimento, finanças e comunicação. Isso significa que a IA é usada em uma variedade de tarefas para as quais o nível de expertise do usuário varia muito: alguém pode ser especialista no serviço que vende e leigo nas obrigações fiscais do negócio. Pela lógica do modelo apresentado na Seção 3, a mesma ferramenta pode gerar economia líquida elevada em uma tarefa e negativa em outra, na mesma semana e para a mesma pessoa.

A terceira é a **internalização integral dos custos de erro**. Quando um erro chega ao cliente, o custo — retrabalho, desconto, reembolso, perda de confiança — recai diretamente sobre o negócio e, em geral, sobre a reputação pessoal de quem o conduz. Além disso, como muitos pequenos negócios não têm como objetivo principal o crescimento, mas a autonomia, a flexibilidade ou outros benefícios não pecuniários (Hurst & Pugsley, 2011), o tempo liberado pode ter valor próprio, independentemente de ser convertido em mais produção. Isso torna a medida de economia líquida de tempo — e não apenas de volume produzido — particularmente adequada a esse público.

Essas características não implicam que a IA seja menos útil para o empreendedor individual; implicam que a forma de avaliar sua utilidade precisa considerar custos que, em organizações maiores, são diluídos ou invisíveis para quem usa a ferramenta.

2.7 Síntese: o que a literatura deixa em aberto

A literatura revisada permite três conclusões provisórias. Primeiro, a IAG pode gerar ganhos expressivos quando a tarefa é bem definida, o produto é fácil de avaliar e o usuário dispõe de conhecimento suficiente para julgá-lo. Segundo, em uso real e em tarefas abertas, os ganhos medidos ou autorrelatados tendem a ser menores, e há evidência de que podem ser nulos ou negativos em contextos específicos. Terceiro, uma parte relevante do trabalho criado

pela IA — especificar, verificar, corrigir, lidar com erros — raramente é medida de forma separada, e a percepção dos usuários sobre seus ganhos pode divergir da medição objetiva.

Não encontramos, na literatura revisada, um modelo que organize esses custos de maneira explícita para o caso do trabalhador ou empreendedor individual, nem estudos que meçam separadamente cada componente em tarefas reais de pequenos negócios; o estudo com empreendedores de Otis et al. (2026) mede resultados do negócio, não a composição do tempo. Essa é a lacuna que o modelo da Seção 3 pretende endereçar.

3 Modelo conceitual

3.1 Unidade de análise e definição

A unidade de análise do modelo é a **tarefa** realizada por um empreendedor individual, entendida como uma unidade de trabalho com um produto identificável e um padrão mínimo de qualidade para ser considerada concluída: uma proposta comercial enviada, uma página de vendas publicada, uma planilha de fluxo de caixa utilizável, uma resposta a um cliente, uma rotina de automação funcionando. A exigência de um padrão de qualidade é essencial: comparar o tempo de duas formas de executar uma tarefa só faz sentido se ambas chegam a um produto de qualidade equivalente.

Definimos **produtividade bruta da IA** em uma tarefa como a diferença entre o tempo que o empreendedor levaria para concluir a tarefa sem IA e o tempo de interação necessário para obter o produto gerado com IA — o contraste entre "fazer do zero" e "receber pronto", que é como o ganho costuma ser percebido. Definimos **produtividade líquida da IA** como a diferença entre o tempo total necessário para concluir a tarefa no padrão de qualidade exigido sem IA e o tempo total necessário com IA, descontado o custo esperado dos erros que permanecem no produto entregue.

3.2 Decomposição contábil

Seja T_0 o tempo necessário para concluir a tarefa sem IA no padrão de qualidade exigido. Com IA, o tempo total T_1 decompõe-se em cinco componentes:

- $\mathbf{T_e}$ **— especificação:** tempo para definir o que se quer, reunir o contexto e redigir as instruções (custo de prevenção, na terminologia de custos da qualidade).
- $\mathbf{T_i}$ **— interação e geração:** tempo de envio, espera, leitura das respostas e iterações de ajuste das instruções. Quando a espera é ocupada com outra atividade — o que se torna

mais comum com agentes que executam tarefas longas —, ela deixa de ser custo da tarefa, mas a retomada passa a alimentar T_s (J. Becker et al., 2026).

- $\mathbf{T_v}$ **— verificação:** tempo para conferir fatos, cálculos, coerência, tom, adequação ao cliente e conformidade com restrições (custo de avaliação).
- $\mathbf{T_r}$ **— correção e retrabalho interno:** tempo para corrigir o que a verificação encontrou ou para refazer, com ou sem IA, partes inadequadas (custo de falha interna).
- $\mathbf{T_s}$ **— troca de contexto:** tempo e perda de desempenho associados à interrupção e retomada do trabalho em torno dos ciclos de interação (Leroy, 2009; Monsell, 2003).

Além disso, o produto entregue pode conter erros que a verificação não detectou. Seja p a probabilidade de que a saída da IA contenha um erro relevante, d a probabilidade de que a verificação o detecte e C o custo, em tempo equivalente, de um erro que chega ao cliente ou ao mercado (retrabalho externo, devoluções, perda de confiança). O custo esperado do erro residual é então $E = p \cdot (1 - d) \cdot C$ (custo de falha externa). Para comparação justa, o mesmo raciocínio aplica-se ao trabalho sem IA, com sua própria probabilidade de erro p_0, probabilidade de detecção d_0 e custo esperado $E_0 = p_0 \cdot (1 - d_0) \cdot C$.

A **economia líquida de tempo** em uma tarefa é, assim:

$$\Delta L = (T_0 + E_0) - (T_e + T_i + T_v + T_r + T_s + E)$$

E o **valor líquido** da economia, para o empreendedor, é ΔL multiplicado pelo custo de oportunidade de seu tempo (G. S. Becker, 1965), acrescido de eventuais diferenças de qualidade acima do padrão mínimo que tenham valor para o cliente.

A **economia bruta** usualmente percebida ou relatada aproxima-se, por outro lado, de $T_0 - T_i$: o contraste entre "fazer do zero" e "receber pronto". A diferença entre as duas medidas — o **hiato bruto-líquido** — é igual a $T_e + T_v + T_r + T_s + (E - E_0)$. O modelo não supõe que esse hiato seja grande ou pequeno; ele torna explícito o que precisa ser medido para sabê-lo.

A probabilidade de detecção não é um parâmetro fixo: ela depende do esforço de verificação, $d = d(T_v)$, crescente e com retornos decrescentes. Disso decorre um nível ótimo de verificação para cada tarefa: vale a pena ampliar T_v enquanto a redução marginal do custo esperado do erro, $p \cdot C \cdot d'(T_v)$, superar o tempo adicional gasto. O ótimo é mais alto quando o erro é provável (p alto) ou caro (C alto). Quando a verificação é pouco eficaz (d' baixo), como tende a ocorrer fora da área de domínio do usuário, ampliar T_v rende pouco, e a

alternativa racional é mudar o método de verificação ou não usar a IA naquela tarefa. Verificação insuficiente e verificação excessiva reduzem, ambas, a economia líquida.

O Quadro 1 sintetiza os componentes, sua correspondência com a tradição de custos da qualidade e a evidência que os ilustra.

Quadro 1. *Componentes da produtividade líquida da IA em uma tarefa*

Componente	Descrição	Categoria de custo da qualidade	Evidência ilustrativa
T_e Especificação	Definir objetivo, contexto, critérios e instruções	Prevenção	Tankelevitch et al. (2024)
T_i Interação	Enviar, esperar, ler, iterar	— (custo operacional)	Barke et al. (2023)
T_v Verificação	Conferir fatos, cálculos, coerência, conformidade	Avaliação	Mozannar et al. (2024)
T_r Retrabalho interno	Corrigir ou refazer antes da entrega	Falha interna	Vaithilingam et al. (2022)
T_s Troca de contexto	Interrupção e retomada; atenção residual	— (custo cognitivo)	Leroy (2009); Mark et al. (2008)
E Erro residual	$p \cdot (1 - d) \cdot C$: erro que chega ao cliente	Falha externa	Magesh et al. (2025); Perry et al. (2023)

Fonte: elaborado pela autora com base em Schiffauerova e Thomson (2006) e na literatura revisada.

3.3 Condições de contorno

Quatro características moderam a magnitude dos componentes. Duas outras — o grau de rotina da tarefa e o tempo de uso da ferramenta — são tratadas, respectivamente, na Proposição 7 e na Seção 3.4.

A **assimetria entre gerar e verificar** é a primeira e, a nosso ver, a mais importante. Há produtos difíceis de gerar e fáceis de verificar — por exemplo, uma lista de títulos alternativos para um anúncio, cuja adequação se avalia em segundos. Há produtos fáceis de gerar e difíceis de verificar — por exemplo, um resumo de legislação tributária aplicável ao negócio, uma projeção financeira ou um contrato. Quanto maior o custo relativo de verificação, maior T_v, menor d para um mesmo esforço e maior o risco de erro residual.

A **expertise do empreendedor no domínio da tarefa** afeta T_0, T_v e d simultaneamente. Quem domina a tarefa tem T_0 baixo — restando pouco a economizar —, mas verifica com rapidez e alta probabilidade de detecção. Quem não domina tem T_0 alto — muito a economizar —, mas verifica com dificuldade e baixa probabilidade de detecção. Essa dupla face ajuda a interpretar por que ganhos maiores entre os menos experientes (Brynjolfsson et al., 2025; Cui et al., 2026) podem coexistir com riscos maiores de erros não detectados (Perry et al., 2023).

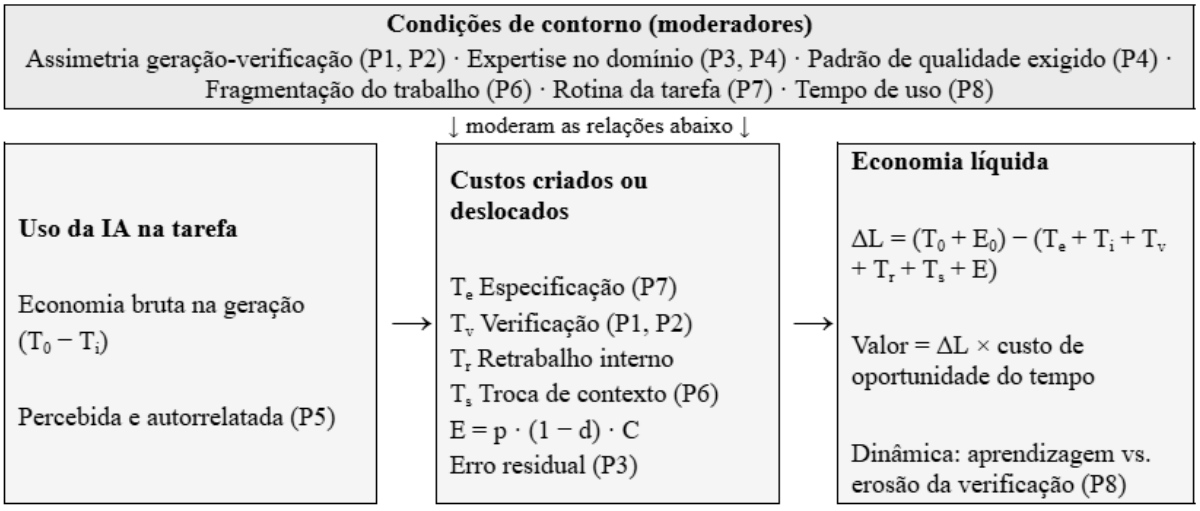
O **padrão de qualidade exigido** determina quanto ajuste é necessário para que o produto gerado seja aceitável. Em contextos com padrões elevados e implícitos — como os projetos maduros estudados por J. Becker et al. (2025) —, o produto gerado pode exigir tanto ajuste que o tempo total aumenta.

A **fragmentação do trabalho** determina T_s. Tarefas executadas em blocos contínuos sofrem menos com trocas de contexto do que tarefas cujo ciclo de interação se intercala com outras atividades.

3.4 Dinâmica ao longo do tempo

O modelo também tem uma dimensão temporal. No início do uso, o empreendedor investe em aprender onde a ferramenta funciona e como especificar tarefas, o que eleva T_e e T_i — análogo, no nível individual, aos investimentos complementares da curva J (Brynjolfsson et al., 2021). Com a experiência, esses custos tendem a cair, e a seleção de tarefas melhora. Em sentido contrário, as ironias da automação (Bainbridge, 1983) sugerem que a capacidade de verificação pode se degradar quando o empreendedor deixa de executar a tarefa por conta própria, reduzindo *d* ao longo do tempo (Bastani et al., 2025; Lee et al., 2025). A Figura 1 representa o modelo.

Figura 1. *Modelo conceitual da produtividade líquida da IA no empreendedorismo individual*



Fonte: elaborado pela autora. As setas indicam relações propostas, não relações testadas. P1 a P8 referem-se às proposições da Seção 4.

4 Proposições

As proposições a seguir decorrem do modelo e da literatura revisada. Elas são formuladas em termos comparativos e com condições explícitas, para que possam ser testadas e refutadas.

4.1 O hiato bruto-líquido

Os ganhos de velocidade na geração são bem documentados, mas a verificação consome parcela relevante do tempo de interação (Mozannar et al., 2024), a correção de saídas inadequadas pode ser trabalhosa (Vaithilingam et al., 2022) e os erros residuais têm custo (Magesh et al., 2025; Pearce et al., 2022). Estudos com economias autorrelatadas modestas em uso real (Bick et al., 2026; Humlum & Vestergaard, 2025) são compatíveis com um hiato relevante entre o ganho bruto e o líquido, embora não o demonstrem.

> **Proposição 1.** Para empreendedores individuais, a economia de tempo medida apenas na etapa de geração superestima sistematicamente a economia líquida; o hiato bruto-líquido é composto principalmente pelos custos de verificação e retrabalho e cresce com o custo relativo de verificação do tipo de produto.

4.2 A assimetria geração-verificação

Se o custo de verificar um produto é baixo em relação ao custo de produzi-lo, a IA pode economizar quase todo o tempo de produção sem criar muito trabalho novo. Se o custo de verificar se aproxima do custo de produzir — porque verificar exige refazer o raciocínio, conferir fontes ou reconstruir cálculos —, a economia líquida tende a desaparecer.

> **Proposição 2.** A economia líquida de tempo da IA é tanto maior quanto menor for a razão entre o custo de verificar o produto e o custo de produzi-lo sem IA; em tarefas em que essa razão se aproxima de um — ou de um limiar inferior a um, uma vez somados os custos de especificação, interação e troca de contexto —, a economia líquida tende a zero ou a valores negativos.

4.3 Expertise e localização dos custos

Os menos experientes tendem a ganhar mais em tarefas estruturadas (Brynjolfsson et al., 2025; Cui et al., 2026), mas também tendem a detectar menos erros e a confiar mais em produtos incorretos (Perry et al., 2023). Entre os muito experientes, que já executam a tarefa com rapidez e padrões elevados, o ganho pode ser pequeno ou negativo (J. Becker et al., 2025). Em pequenos negócios, Otis et al. (2026) acrescentam uma nuance: os empreendedores de menor desempenho perderam com a IA não por receberem conselhos piores, mas por

escolherem pior quais implementar — isto é, por menor capacidade de julgar o que foi gerado.

> **Proposição 3.** A expertise do empreendedor no domínio da tarefa desloca a composição dos custos: empreendedores com menor expertise obtêm maiores economias de tempo visíveis, mas incorrem em maior custo esperado de erro residual; empreendedores com maior expertise têm menores economias visíveis, mas menor custo de erro residual.

> **Proposição 4.** Em tarefas nas quais o empreendedor tem alta expertise e um padrão de qualidade elevado e específico, a economia líquida da IA tende a ser pequena e pode ser negativa, porque o ajuste necessário para atingir o padrão compensa o tempo poupado na geração.

4.4 A lacuna de percepção

A falácia do planejamento (Buehler et al., 1994), a associação entre fluidez da interação e produtividade percebida (Ziegler et al., 2024) e a discrepância entre ganho estimado e medido documentada por J. Becker et al. (2025) sugerem que as autoavaliações de economia de tempo são vulneráveis a vieses. Esses vieses tenderiam a ser maiores quando os custos de verificação e retrabalho estão dispersos no tempo e não são registrados como parte da tarefa.

> **Proposição 5.** A economia de tempo autorrelatada por empreendedores individuais tende a exceder a economia líquida medida; essa diferença é maior quando a verificação e o retrabalho ocorrem de forma fragmentada ou depois da entrega, em vez de concentrados na execução da tarefa.

4.5 Fragmentação e troca de contexto

O empreendedor individual acumula funções — produção, venda, atendimento, finanças —, e os ciclos de pedir, esperar, conferir e corrigir podem se intercalar com outras atividades. A literatura sobre troca de tarefas e atenção residual (Leroy, 2009; Monsell, 2003) e sobre interrupções (Mark et al., 2008) sugere custos que não aparecem nas medidas de duração da tarefa, mas afetam o desempenho e o esforço.

> **Proposição 6.** Quanto mais fragmentado for o ciclo de uso da IA — isto é, quanto mais a interação, a espera e a verificação se intercalarem com outras atividades —, menor será a economia líquida, por efeito dos custos de troca de contexto, ainda que o tempo de geração permaneça constante.

4.6 Especificação como prevenção

Na lógica de custos da qualidade, investir em prevenção reduz custos de falha (Schiffauerova & Thomson, 2006), e corrigir problemas nas fases iniciais tende a custar menos do que depois da entrega (Boehm & Basili, 2001). No uso da IA, a especificação de objetivos, contexto e critérios de aceitação é a principal forma de prevenção, e a explicitação dos objetivos é uma demanda metacognitiva central da interação com modelos de linguagem (Tankelevitch et al., 2024; ver também Acar, 2023).

> **Proposição 7.** Em tarefas não rotineiras, o tempo investido na especificação prévia da tarefa — objetivo, contexto, restrições e critérios de aceitação — reduz os custos de verificação, retrabalho e erro residual em magnitude superior ao seu próprio custo; em tarefas rotineiras e fáceis de verificar, esse benefício líquido tende a ser pequeno.

4.7 Dinâmica temporal: aprendizagem e erosão da verificação

No nível individual, a adoção da IA envolve aprendizagem que inicialmente reduz a produtividade medida e depois a aumenta (Brynjolfsson et al., 2021). Ao mesmo tempo, a transferência da execução para a IA pode erodir a capacidade de verificar (Bainbridge, 1983; Bastani et al., 2025; Lee et al., 2025).

> **Proposição 8.** Ao longo do tempo de uso, a economia líquida da IA segue duas dinâmicas opostas: aumenta à medida que o empreendedor aprende a selecionar tarefas e a especificá-las, e diminui se a menor prática na execução reduzir sua capacidade de detectar erros; o saldo depende de o empreendedor manter práticas deliberadas de verificação.

4.8 Relações entre as proposições

As proposições não são independentes. As Proposições 1 e 2 descrevem o núcleo do modelo: o hiato bruto-líquido existe e é governado principalmente pela assimetria entre gerar e verificar. As Proposições 3 e 4 especificam como a expertise do usuário redistribui esse hiato entre custos visíveis (tempo de verificação e ajuste) e custos invisíveis (erros residuais). A Proposição 5 trata de um problema de mensuração que atravessa todas as demais: se as autoavaliações são enviesadas, estudos baseados apenas em relatos tenderão a superestimar os ganhos, o que reforça a necessidade de medição objetiva por componente. A Proposição 6 introduz um custo que é específico da organização do trabalho, e não da tarefa. As Proposições 7 e 8 tratam do que o empreendedor pode fazer — investir em prevenção e preservar a capacidade de verificar — e de como o saldo evolui com o tempo.

Uma implicação dessa estrutura é que a mesma ferramenta, usada pela mesma pessoa, pode produzir economia líquida positiva em uma tarefa e negativa em outra. Isso desloca a pergunta prática de "a IA economiza tempo?" para "em quais tarefas, com que forma de uso e para quem a IA economiza tempo depois de contabilizados todos os custos?".

Quadro 2. *Síntese das proposições*

Prop.	Relação proposta	Variáveis-chave	Principais bases
P1	Economia bruta > economia líquida	Custo relativo de verificação	Mozannar et al. (2024); Bick et al. (2026)
P2	Razão verificação/produção ↓ → economia líquida ↑	Assimetria geração-verificação	Pearce et al. (2022); Magesh et al. (2025)
P3	Expertise desloca custos visíveis e residuais	Expertise; erro residual	Brynjolfsson et al. (2025); Perry et al. (2023); Otis et al. (2026)
P4	Alta expertise + padrão elevado → economia pequena ou negativa	Padrão de qualidade	J. Becker et al. (2025)
P5	Economia autorrelatada > economia medida	Dispersão do retrabalho	Buehler et al. (1994); Ziegler et al. (2024)
P6	Fragmentação ↑ → economia líquida ↓	Troca de contexto	Leroy (2009); Monsell (2003)
P7	Especificação prévia reduz custos de falha	Rotina da tarefa	Boehm e Basili (2001); Tankelevitch et al. (2024)
P8	Aprendizagem ↑ e erosão da verificação ↓	Tempo de uso	Brynjolfsson et al. (2021); Bainbridge (1983)

Fonte: elaborado pela autora.

5 Implicações para empreendedores individuais

As implicações a seguir decorrem do modelo e devem ser lidas como hipóteses práticas, não como recomendações comprovadas.

A primeira implicação é **medir a tarefa inteira, e não apenas a geração**. Um registro simples, mantido por algumas semanas, do tempo gasto em cada fase — especificar, gerar, verificar, corrigir — e das correções feitas depois da entrega permite estimar a economia líquida em vez de confiar na impressão de rapidez. A Proposição 5 sugere que a impressão tende a ser otimista.

A segunda implicação é **escolher tarefas pela assimetria geração-verificação**. Tarefas cujo produto é fácil de avaliar — alternativas de redação, estruturas de documentos, primeiros rascunhos que serão inteiramente reescritos, listas de ideias — são candidatas naturais. Tarefas cujo produto exige verificação extensa — números, normas, fatos sobre terceiros, compromissos contratuais — requerem um protocolo de verificação explícito ou podem não compensar (Proposição 2).

A terceira implicação é **tratar a verificação como parte da tarefa, e não como etapa opcional**. Na terminologia dos custos da qualidade, economizar em avaliação desloca custo para a falha externa, que, para quem trabalha sozinho, recai sobre a própria reputação. Um critério prático derivado do modelo é proporcional: quanto maior o custo de um erro chegar ao cliente (C), maior deve ser o esforço de verificação. Quando a pouca familiaridade com o tema torna a verificação pouco eficaz, porém, reler mais vezes não resolve: o modelo indica verificar contra fontes externas, recorrer a quem domina o tema ou não delegar a tarefa à IA (Seção 3.2).

A quarta implicação é **reduzir a fragmentação**. Agrupar tarefas de IA em blocos, em vez de intercalar pedidos e verificações com outras atividades ao longo do dia, pode reduzir custos de troca de contexto (Proposição 6).

A quinta implicação é **investir em especificação nas tarefas não rotineiras e reutilizá-la nas rotineiras**. A explicitação do objetivo e dos critérios de aceitação antes de pedir o produto funciona como prevenção (Proposição 7); para tarefas recorrentes, essa especificação pode ser feita uma vez e reaproveitada. O bloco Questionar do método AI Thinking, que motivou a autora a investigar o tema, enfatiza essa explicitação; o modelo aqui apresentado oferece uma forma de avaliar empiricamente se ela compensa.

A sexta implicação é **preservar a capacidade de executar sem a IA** as tarefas críticas do negócio, como salvaguarda da capacidade de verificar (Proposição 8).

6 Limitações

A primeira limitação é a ausência de dados empíricos. O modelo é contábil e conceitual; as proposições derivam de estudos que, em sua maioria, não mediram separadamente os componentes aqui propostos nem estudaram empreendedores individuais.

A segunda limitação refere-se à generalização das evidências. Grande parte dos estudos citados envolve programadores, consultores, profissionais de grandes empresas ou trabalhadores de países de alta renda. A principal exceção com empreendedores (Otis et al., 2026) foi conduzida no Quênia e mediu resultados do negócio, não tempo. A programação está sobrerrepresentada, porque nela o tempo e os defeitos são mais facilmente observáveis. Não se sabe em que medida os padrões observados se repetem em tarefas típicas de pequenos negócios, como marketing, atendimento, gestão financeira e produção de conteúdo. Além disso, as ferramentas evoluem rapidamente; os próprios autores do estudo com desenvolvedores experientes alertaram que seus resultados retratam um momento específico e

relataram dificuldades crescentes de mensuração com ferramentas posteriores (J. Becker et al., 2025, 2026).

A terceira limitação é de mensuração. O modelo supõe que a qualidade possa ser fixada em um padrão comparável entre as condições com e sem IA, o que é difícil em tarefas criativas ou abertas. O custo de erros residuais (C) é particularmente difícil de estimar, pois inclui efeitos reputacionais e perdas de oportunidade que podem se manifestar muito depois. A separação entre tempo de verificação e tempo de retrabalho também é, na prática, imperfeita.

A quarta limitação é conceitual. O modelo trata o tempo como principal insumo e ignora outros efeitos possivelmente relevantes da IA, como a ampliação do conjunto de tarefas que o empreendedor consegue executar, mudanças na qualidade acima do padrão mínimo, efeitos sobre bem-estar e motivação e a possibilidade de que o tempo economizado não seja convertido em atividades de maior valor. Um empreendedor pode, racionalmente, aceitar economia líquida de tempo nula em troca de um produto melhor ou de uma experiência de trabalho mais agradável.

A quinta limitação refere-se à posição da autora, que desenvolve um método de formação no tema. Para mitigar esse risco, o método não é usado como evidência, as proposições são refutáveis e o desenho de estudo proposto prevê pré-registro, avaliação cega e análise independente (Seção 7).

7 Agenda de pesquisa e desenho de estudo proposto

7.1 Questões de pesquisa

A agenda decorrente do modelo organiza-se em torno de quatro questões. (Q1) Qual é a magnitude do hiato entre economia bruta e economia líquida de tempo da IA em tarefas reais de empreendedores individuais? (Q2) Quais componentes — especificação, interação, verificação, retrabalho, troca de contexto, erro residual — explicam esse hiato e como variam conforme a assimetria geração-verificação, a expertise e o padrão de qualidade? (Q3) Em que medida as autoavaliações de economia de tempo correspondem à medição objetiva? (Q4) Um protocolo estruturado de especificação e verificação altera a economia líquida?

Essas questões admitem estratégias complementares. Estudos de diário e de registro de tempo podem descrever a distribuição real dos componentes em pequenos negócios. Estudos observacionais com gravação de tela e codificação de comportamentos, na linha de Mozannar

et al. (2024), podem estimar a parcela de tempo dedicada à verificação em tarefas não relacionadas à programação. Experimentos com tarefas padronizadas permitem manipular a assimetria geração-verificação com alto controle. O teste mais exigente, porém, é um experimento com tarefas reais, descrito a seguir.

7.2 Desenho proposto: randomização no nível da tarefa com três condições

Propomos um experimento com empreendedores individuais e proprietários de microempresas sem empregados ou com poucos empregados, inspirado no desenho de J. Becker et al. (2025), em que tarefas reais são definidas antes da randomização. Cada participante submete uma lista de tarefas que precisa realizar em seu negócio nas semanas seguintes, com descrição, padrão de conclusão e estimativa de tempo sem IA. Cada tarefa é então aleatoriamente designada a uma de três condições:

- **Condição A (sem IA):** a tarefa é executada sem ferramentas de IA generativa.
- **Condição B (IA de uso livre):** a tarefa pode ser executada com as ferramentas de IA que o participante desejar, sem orientação adicional.
- **Condição C (IA com protocolo):** a tarefa é executada com IA, seguindo um protocolo breve que exige especificação prévia (objetivo, contexto, restrições e critérios de aceitação) e uma lista de verificação proporcional ao risco antes da entrega.

A comparação B versus A estima o efeito líquido do uso habitual da IA. A comparação C versus B estima a contribuição específica do protocolo de especificação e verificação, testando a Proposição 7. A randomização no nível da tarefa, dentro de cada participante, controla diferenças estáveis entre empreendedores e permite comparar condições no mesmo negócio. Para reduzir efeitos de aprendizagem do protocolo sobre a condição B, recomenda-se que o treinamento no protocolo ocorra apenas depois de uma primeira fase em que só as condições A e B são sorteadas, ou, alternativamente, que a condição C seja comparada em um desenho entre participantes. A escolha deve ser justificada no pré-registro.

Uma dificuldade relevante, apontada pelos próprios autores do estudo que inspira o desenho, é a seleção: participantes podem relutar em realizar tarefas sem IA ou podem submeter preferencialmente tarefas em que esperam ganho (J. Becker et al., 2026). Recomenda-se que o desenho preveja o registro de tarefas recusadas ou retiradas, incentivos que não dependam da condição sorteada e análises de sensibilidade para seleção.

7.3 Hipóteses

As hipóteses confirmatórias, a serem pré-registradas, são:

- **H1 (P1):** Na condição B, a economia de tempo medida até a geração do primeiro produto excede a economia líquida medida até a conclusão da tarefa no padrão exigido.
- **H2 (P2):** A economia líquida na condição B, em relação à condição A, diminui com a razão entre tempo de verificação e tempo de execução sem IA, classificada para cada tarefa por avaliadores independentes antes da execução.
- **H3 (P5):** A economia de tempo autorrelatada pelos participantes ao final do estudo excede a economia líquida medida.
- **H4 (P7):** Em tarefas não rotineiras, a economia líquida na condição C excede a da condição B.

As Proposições 3, 4, 6 e 8 podem ser examinadas como hipóteses secundárias ou exploratórias, explicitamente rotuladas como tal: P3 e P4 por interações entre condição e expertise autodeclarada e avaliada; P6 pela fragmentação registrada nos logs; P8 por tendências ao longo do período e por uma tarefa final sem IA.

7.4 Medidas

Tempo por componente. Gravação de tela com consentimento, complementada por registro de atividades, permite codificar o tempo de especificação, interação, verificação e retrabalho, seguindo a lógica de taxonomias de comportamento como a de Mozannar et al. (2024). Uma amostra das gravações deve ser codificada por dois avaliadores, com confiabilidade reportada.

Troca de contexto. Número de interrupções e de alternâncias entre aplicativos e tarefas durante a execução, extraído dos registros, e esforço percebido ao final de cada tarefa (Mark et al., 2008).

Qualidade do produto. Avaliação cega por especialistas do domínio, com rubrica específica por tipo de tarefa e verificação do atingimento do padrão de conclusão definido antes da randomização. A comparação de tempo deve condicionar-se ao atingimento do padrão.

Erros e retrabalho externo. Acompanhamento de cada produto entregue por um período definido (por exemplo, 30 a 60 dias), registrando correções posteriores, reclamações, devoluções e retrabalho. Quando possível, uma auditoria independente de uma amostra de produtos estima a taxa de erros não detectados.

Percepção. Previsão de economia de tempo antes da tarefa e estimativa posterior, bem como uma estimativa global ao final do estudo, para comparação com a medição objetiva (J. Becker et al., 2025; Buehler et al., 1994).

Características da tarefa e do participante. Razão estimada verificação/execução, rotina ou não rotina, custo potencial de um erro, expertise no domínio e experiência prévia com IA.

7.5 Tamanho amostral e análise

O tamanho amostral deve ser definido por análise de poder a priori e justificado no pré-registro. A título ilustrativo — e sem que isso represente uma estimativa do efeito esperado —, considere a análise principal no nível do participante, comparando a economia líquida média entre duas condições dentro de cada participante. Para detectar uma diferença padronizada dos escores de diferença de $d = 0{,}30$, com poder de 0,80 e $\alpha = 0{,}05$ bicaudal, um teste pareado exige cerca de 90 participantes (cálculo exato com a distribuição t não central). Se as duas comparações confirmatórias principais (B versus A e C versus B) forem corrigidas para multiplicidade com $\alpha = 0{,}025$ cada, o requisito sobe para cerca de 109 participantes. Se a condição C for comparada em desenho entre participantes, a mesma diferença padronizada de $d = 0{,}30$ exigiria cerca de 176 participantes por grupo, com poder de 0,80 e $\alpha = 0{,}05$.

Esses números supõem que cada participante contribua com tarefas suficientes em cada condição para estimar sua média de forma estável. Se a análise for conduzida no nível da tarefa, com tarefas agrupadas em participantes, o tamanho efetivo da amostra deve ser ajustado pelo efeito de desenho, $1 + (m - 1) \cdot ICC$, em que *m* é o número médio de tarefas por participante e ICC é a correlação intraclasse dos resultados; ambos devem ser estimados em estudo piloto, e não presumidos. A esses valores deve-se acrescentar margem para perdas, que tendem a ser relevantes em estudos de campo com pequenos negócios.

Como o tempo de execução tende a ter distribuição assimétrica, recomenda-se especificar no pré-registro a análise em escala logarítmica, que permite interpretar os efeitos como variações percentuais. A análise principal deve usar modelos de efeitos mistos com efeitos aleatórios de participante e controles para a estimativa de tempo prévia e o tipo de tarefa. Hipóteses exploratórias devem ser rotuladas como tal e reportadas integralmente, inclusive quando nulas.

7.6 Ética e transparência

O estudo deve ser submetido a comitê de ética em pesquisa — no Brasil, ao sistema CEP/Conep —, com consentimento informado específico para gravação de tela, proteção de dados de clientes e do negócio e possibilidade de desistência a qualquer momento. Como a condição sem IA e os erros eventualmente introduzidos pela IA podem afetar o negócio, os

participantes devem manter a decisão final sobre o que entregar a seus clientes, e deve haver procedimento para registrar e mitigar efeitos adversos.

Dada a origem do estudo em um método de formação desenvolvido pela autora, recomenda-se que a codificação, a avaliação de qualidade e a análise sejam conduzidas ou auditadas por pesquisadores sem vínculo com o método; que o pré-registro especifique hipóteses, medidas e análises antes da coleta; e que materiais, rubricas e, quando possível, dados anonimizados sejam disponibilizados publicamente.

8 Conclusão

Este ensaio examinou quanto a IA economiza para o empreendedor individual depois da revisão e do retrabalho. Argumentamos que a resposta depende de distinguir produtividade bruta de produtividade líquida e propusemos um modelo que decompõe a segunda em componentes mensuráveis — especificação, interação, verificação, retrabalho, troca de contexto e erro residual —, ancorado na tradição dos custos da qualidade e na pesquisa sobre automação. Derivamos oito proposições e um desenho de estudo capaz de testá-las.

A principal implicação teórica do modelo é que a economia de tempo da IA não é uma propriedade da ferramenta, mas da combinação entre tarefa, usuário e forma de uso. Ela tende a ser alta quando o produto é fácil de verificar e o usuário sabe julgá-lo, e pode ser nula ou negativa quando verificar custa quase tanto quanto produzir ou quando o padrão de qualidade exige ajustes extensos. Para o empreendedor individual, que concentra em si a geração, a verificação e a responsabilidade pelos erros, essa distinção é especialmente consequente.

A implicação metodológica é que medir apenas a velocidade de geração, ou confiar em autoavaliações, tende a produzir estimativas otimistas. Se as proposições forem confirmadas, a avaliação do uso da IA em pequenos negócios deveria incluir sistematicamente os custos de verificação e retrabalho. Se forem refutadas — por exemplo, se esses custos se mostrarem pequenos em tarefas típicas de pequenos negócios —, o resultado será igualmente informativo. Em ambos os casos, a questão merece ser tratada como objeto de investigação empírica, e não como pressuposto.

Declarações

Conflito de interesses. A autora atua como empreendedora e na formação de líderes e empreendedores na área de inteligência artificial e é autora do método AI Thinking, mencionado neste artigo como motivação.

Uso de inteligência artificial. Na preparação deste artigo, a autora utilizou o Claude (Anthropic) para revisão do texto, conferência das referências e verificação do cálculo de poder. A autora revisou integralmente o conteúdo e assume total responsabilidade por ele.

Financiamento. Esta pesquisa não recebeu financiamento específico de agências públicas, comerciais ou sem fins lucrativos.

Disponibilidade de dados. Não se aplica. Este artigo é teórico e não utiliza dados empíricos.

Referências

Acar, O. A. (2023, June 6). AI prompt engineering isn't the future. *Harvard Business Review*. https://hbr.org/2023/06/ai-prompt-engineering-isnt-the-future

Acemoglu, D. (2025). The simple macroeconomics of AI. *Economic Policy, 40*(121), 13–58. https://doi.org/10.1093/epolic/eiae042

Bainbridge, L. (1983). Ironies of automation. *Automatica, 19*(6), 775–779. https://doi.org/10.1016/0005-1098(83)90046-8

Barke, S., James, M. B., & Polikarpova, N. (2023). Grounded Copilot: How programmers interact with code-generating models. *Proceedings of the ACM on Programming Languages, 7*(OOPSLA1), 85–111. https://doi.org/10.1145/3586030

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences, 122*(26), Article e2422633122. https://doi.org/10.1073/pnas.2422633122

Becker, G. S. (1965). A theory of the allocation of time. *The Economic Journal, 75*(299), 493–517. https://doi.org/10.2307/2228949

Becker, J., Rush, N., Barnes, E., & Rein, D. (2025). *Measuring the impact of early-2025 AI on experienced open-source developer productivity* (arXiv:2507.09089). arXiv. https://doi.org/10.48550/arXiv.2507.09089

Becker, J., Rush, N., Cunningham, T., Rein, D., & Mahamud, K. (2026, February 24). *We are changing our developer productivity experiment design*. METR. https://metr.org/blog/2026-02-24-uplift-update/

Bick, A., Blandin, A., & Deming, D. J. (2026). The rapid adoption of generative AI. *Management Science, 72*(10), 8440–8456. https://doi.org/10.1287/mnsc.2025.02523

Boehm, B., & Basili, V. R. (2001). Software defect reduction top 10 list. *Computer, 34*(1), 135–137. https://doi.org/10.1109/2.962984

Brynjolfsson, E. (1993). The productivity paradox of information technology. *Communications of the ACM, 36*(12), 66–77. https://doi.org/10.1145/163298.163309

Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at work. *The Quarterly Journal of Economics, 140*(2), 889–942. https://doi.org/10.1093/qje/qjae044

Brynjolfsson, E., Rock, D., & Syverson, C. (2021). The productivity J-curve: How intangibles complement general purpose technologies. *American Economic Journal: Macroeconomics, 13*(1), 333–372. https://doi.org/10.1257/mac.20180386

Buehler, R., Griffin, D., & Ross, M. (1994). Exploring the "planning fallacy": Why people underestimate their task completion times. *Journal of Personality and Social Psychology, 67*(3), 366–381. https://doi.org/10.1037/0022-3514.67.3.366

Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction, 5*(CSCW1), 1–21. https://doi.org/10.1145/3449287

Cui, K. Z., Demirer, M., Jaffe, S., Musolff, L., Peng, S., & Salz, T. (2026). The effects of generative AI on high-skilled work: Evidence from three field experiments with software developers. *Management Science*. Advance online publication. https://doi.org/10.1287/mnsc.2025.00535

Dell'Acqua, F., Ayoubi, C., Lifshitz, H., Sadun, R., Mollick, E., Mollick, L., Han, Y., Goldman, J., Nair, H., Taub, S., & Lakhani, K. R. (2026). The cybernetic teammate: A field experiment on generative AI and teamwork. *Organization Science, 37*(4), 1217–1242. https://doi.org/10.1287/orsc.2025.20702

Dell'Acqua, F., McFowland, E., Mollick, E., Lifshitz, H., Kellogg, K. C., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2026). Navigating the jagged technological frontier: Field experimental evidence of the effects of artificial intelligence on knowledge worker productivity and quality. *Organization Science, 37*(2), 403–423. https://doi.org/10.1287/orsc.2025.21838

Dillon, E., Jaffe, S., Immorlica, N., & Stanton, C. (2025). *Shifting work patterns with generative AI* (NBER Working Paper No. 33795). National Bureau of Economic Research. https://doi.org/10.3386/w33795

Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation bias: A systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association, 19*(1), 121–127. https://doi.org/10.1136/amiajnl-2011-000089

Humlum, A., & Vestergaard, E. (2025). *Large language models, small labor market effects* (NBER Working Paper No. 33777). National Bureau of Economic Research. https://doi.org/10.3386/w33777 [Revisões posteriores foram publicadas com o título *Still waters, rapid currents: Early labor market transformation under generative AI*.]

Hurst, E., & Pugsley, B. W. (2011). What do small businesses do? *Brookings Papers on Economic Activity, 2011*(2), 73–118. https://doi.org/10.1353/eca.2011.0017

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys, 55*(12), 1–38. https://doi.org/10.1145/3571730

Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (pp. 1–22). Association for Computing Machinery. https://doi.org/10.1145/3706598.3713778

Leroy, S. (2009). Why is it so hard to do my work? The challenge of attention residue when switching between work tasks. *Organizational Behavior and Human Decision Processes, 109*(2), 168–181. https://doi.org/10.1016/j.obhdp.2009.04.002

Magesh, V., Surani, F., Dahl, M., Suzgun, M., Manning, C. D., & Ho, D. E. (2025). Hallucination-free? Assessing the reliability of leading AI legal research tools. *Journal of Empirical Legal Studies, 22*(2), 216–242. https://doi.org/10.1111/jels.12413

Mark, G., Gudith, D., & Klocke, U. (2008). The cost of interrupted work: More speed and stress. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 107–110). Association for Computing Machinery. https://doi.org/10.1145/1357054.1357072

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences, 7*(3), 134–140. https://doi.org/10.1016/S1364-6613(03)00028-7

Mozannar, H., Bansal, G., Fourney, A., & Horvitz, E. (2024). Reading between the lines: Modeling user behavior and costs in AI-assisted programming. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (pp. 1–16). Association for Computing Machinery. https://doi.org/10.1145/3613904.3641936

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science, 381*(6654), 187–192. https://doi.org/10.1126/science.adh2586

Otis, N. G., Clarke, R., Delecourt, S., Holtz, D., & Koning, R. (2026). The uneven impact of generative AI on entrepreneurial performance: Evidence from a field experiment in Kenya. *Management Science*. Advance online publication. https://doi.org/10.1287/mnsc.2024.06909

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors, 52*(3), 381–410. https://doi.org/10.1177/0018720810376055

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*(2), 230–253. https://doi.org/10.1518/001872097778543886

Pearce, H., Ahmad, B., Tan, B., Dolan-Gavitt, B., & Karri, R. (2022). Asleep at the keyboard? Assessing the security of GitHub Copilot's code contributions. In *2022 IEEE*

Symposium on Security and Privacy (SP) (pp. 754–768). IEEE. https://doi.org/10.1109/SP46214.2022.9833571

Peng, S., Kalliamvakou, E., Cihon, P., & Demirer, M. (2023). *The impact of AI on developer productivity: Evidence from GitHub Copilot* (arXiv:2302.06590). arXiv. https://doi.org/10.48550/arXiv.2302.06590

Perry, N., Srivastava, M., Kumar, D., & Boneh, D. (2023). Do users write more insecure code with AI assistants? In *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security* (pp. 2785–2799). Association for Computing Machinery. https://doi.org/10.1145/3576915.3623157

Schiffauerova, A., & Thomson, V. (2006). A review of research on cost of quality models and best practices. *International Journal of Quality & Reliability Management, 23*(6), 647–669. https://doi.org/10.1108/02656710610672470

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Syverson, C. (2011). What determines productivity? *Journal of Economic Literature, 49*(2), 326–365. https://doi.org/10.1257/jel.49.2.326

Tankelevitch, L., Kewenig, V., Simkute, A., Scott, A. E., Sarkar, A., Sellen, A., & Rintel, S. (2024). The metacognitive demands and opportunities of generative AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (pp. 1–24). Association for Computing Machinery. https://doi.org/10.1145/3613904.3642902

Vaithilingam, P., Zhang, T., & Glassman, E. L. (2022). Expectation vs. experience: Evaluating the usability of code generation tools powered by large language models. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1–7). Association for Computing Machinery. https://doi.org/10.1145/3491101.3519665

Ziegler, A., Kalliamvakou, E., Li, X. A., Rice, A., Rifkin, D., Simister, S., Sittampalam, G., & Aftandilian, E. (2024). Measuring GitHub Copilot's impact on productivity. *Communications of the ACM, 67*(3), 54–63. https://doi.org/10.1145/3633453