ChatGPT

AIは専門家の技能を高めるのか、それとも代替するのか

特許弁護士133人・3か月RCTを中心とした原典検証と関連研究レビュー

エグゼクティブサマリ

本件の中核研究は、David Autor らによる NBER Working Paper “Does AI Assistance Enhance or Erode Expertise? Evidence from a Three-Month Field Experiment in Patent Drafting” である。研究は事前登録された3か月のランダム化比較試験（RCT）で、米国の知的財産法律事務所11社に所属する現役特許弁護士133人を、Google Labs製の特許ドラフティング支援AIにアクセスできる群90人と対照群43人に割り付けた。AI群は日常業務でもツールを利用し、約10日後と90日後に特許ドラフティング課題を実施した。90日後には両群ともAIを使わず、既存の特許出願書類を修正する redlining 課題を行った。評価は独立した特許弁護士が処置群を知らない状態で行った。 [1]

主要結果は明確である。AIへのアクセスは、AIを使えるドラフティング課題の品質を10日時点で **+0.34 SD（p=0.03）**、90日時点で **+0.38 SD（p=0.01）** 改善した。90日後のAIなしredliningでも全体では **+0.32 SD（p=0.04）** だったが、この効果はほぼ全面的に経験7年以上のシニアに集中し、シニアでは **+0.45 SD（p=0.02）**、ジュニアでは **−0.03 SD（p=0.89）** だった。したがって、「AIで成果物の品質が上がる」ことと「AIなしでも本人の専門技能が向上する」ことは別の現象である。 [2]

一方で、「AIは若手の技能を侵食する」とまで結論するのは証拠を超える。ジュニアのAIなし課題の平均成績は対照群より低くなってはいない。むしろ分布が二極化し、低得点者と比較的良い得点者が増え、中間層が減った。シニアとジュニアのAIなし効果差は統計的に有意（p=0.04）だったが、ジュニア完遂者は少なく、最終redliningまで完了したのは全体91人、そのうちジュニアは24人にすぎない。 [3]

本研究の最も重要な弱点は、①治療前の「AIなし能力」のベースライン測定がない、②最終課題でAI不使用を完全には検証できない、③133人中91人しか最終課題を完了していない、④参加事務所がGoogleと継続的に取引する11社に限定される、⑤Googleがツールを開発・研究費を負担し、7著者中6人がGoogle関係者である、⑥3か月は専門技能形成として短い、という点である。AI不使用違反の可能性がある15件を除外すると、シニアの効果は約0.39 SDまで縮み、p=0.06となり、全体効果は0.25 SD、p=0.14となった。したがって、最も堅い結論は「シニアに有望な技能移転のシグナルがある」であり、「技能向上が確立した」ではない。 [4]

PsyPostの記事は元論文の主要結果と限界を概ね忠実に伝えている。ただしニュース記事では、「ジュニアはAIを使っている間に最も恩恵を受けたのに、AIなしでは伸びなかった」という物語が強調される一方、元研究上はジュニア対シニアのドラフティング効果差自体は10日・90日のいずれも統計的有意ではない。したがって、「若手ほど確実にAI恩恵が大きい」と断定すべきではない。 [5]

指定された2026年11月8日付TechFeed記事については、公開検索で当該RCTを扱うページを一意に特定できなかった。TechFeedはPsyPost等の海外記事を日本語で紹介する形式を日常的に用いる媒体であるが、本分析ではTechFeedの二次記述を証拠源とはせず、PsyPost、NBER、AEA RCT Registry、Google Research、著者によるCEPR解説を相互照合している。 [6]

方法とデータ

原研究の設計

研究対象候補は、Googleと定常的かつ非独占的に特許業務を行う24の知財法律事務所だった。そのうち11事務所が参加し、1事務所については2コホートを時期をずらして扱ったため、解析上は12のfirm-groupとなった。事務所側から156人が推薦され、133人がonboardingを完了してランダム化された。割付は各事務所内で約2対1、AI群90人・対照群43人であり、経験年数も考慮した層別化が行われたと報告されている。 [4]

事前登録は AEA RCT Registry の **AEARCTR-0015823** で、2025年5月21日に公開された。登録時の主要アウトカムは、独立評価者による課題品質、作業時間、自己申告の品質・時間変化、業務満足度などだった。対照群は通常の業務プロセスを用い、介入群は専用AIツールと特許ドラフティングにおけるAI利用訓練を受ける設計である。 [7]

項目	介入群	対照群	原典上の注意点
ランダム化人数	90	43	計133人
国	米国	米国	11知財事務所
日常業務で専用AI	あり	なし	対照群は通常ツール
10日課題	AI利用可	AIなし	特許ドラフト
90日課題	AI利用可	AIなし	より難しいドラフト
90日redlining	AI禁止	AI禁止	専門判断の移転評価
評価者	独立特許弁護士	同左	処置割付にblind
最終課題完遂	合計91人	合計91人内	群別完遂数は本文表確認が必要
若手定義	<7年	<7年	シニアは≥7年

[8]

被験者属性

原典および事前登録から確実に確認できるのは、参加者が米国11の知財法律事務所に勤務する現役のpatent lawyersであること、分析上「7年未満」をジュニア、「7年以上」をシニアと定義したことである。 [9]

平均年齢、年齢分布、性別、具体的な平均経験年数、各州・各事務所の所在地、パートナー／アソシエイト比率については、公開検索で確認できた原典情報では明確な記載を確認できなかったため、本レポートでは「未記載」と扱う。「米国人」という意味ではなく、「米国の事務所で勤務する弁護士」である点にも注意が必要である。 [10]

AI介入の内容

介入ツールは当時未公開だったGoogle Labsのプロフェッショナル・ライティング支援ツール **InFlow** とされ、後にGemini Notebookに組み込まれたとGoogleは説明している。利用者が文脈、文体、執筆目標を与え、プロンプトによる文章修正、複数の仮想編集者視点によるフィードバック、文書内不整合チェックなどを行えた。治療群には特許業務向け設定、プロンプトライブラリ、AIドラフティングの典型的な落とし穴に関するガイダンスも提供された。 [11]

ただし、**基盤LLMの正確なモデル名・バージョン、temperature、system prompt、全プロンプトの固定性について、公開された要約資料からは特定できない**。Google自身も研究対象が2024〜2025年頃のAI能力を反映すると説明しており、2026年時点のモデルとの直接的同一視はできない。 12

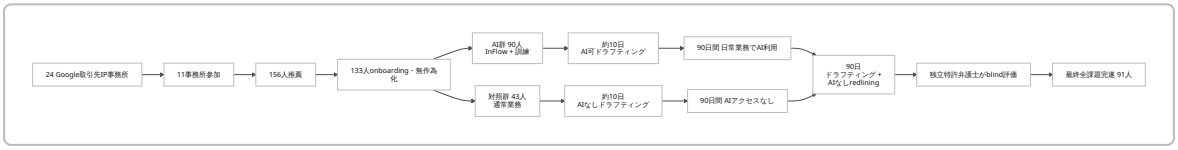
研究期間中にツール自体が更新され、自動redlining機能が追加されたとの記述がある。この点は「固定された単一介入」を評価する通常の薬剤RCTとは異なり、介入内容の時間変化を意味する。 13

評価課題と追跡

約10日後の課題では、主請求項、図面、発明者ノートなどから5〜7本の従属請求項と詳細説明を作るよう求められ、時間上限は約2時間、文章量は約1,000語とされた。90日後には別の、より難しいドラフティング課題を実施した。 14

同じ90日時点で、全員がAIを使わず、既存のAI生成特許草稿を修正するredlining課題を行った。これは新規文章生成よりも、問題発見、権利範囲の判断、不要な限定の削除、戦略的修正といった専門判断を要求するよう設計された。 15

評価軸は、**enforceability、technical accuracy、strategic ambiguity、completeness/alignment、clarity** の5領域で、1〜5段階評価が中心だった。独立法律事務所の特許弁護士が割付を知らずに採点したため、アウトカム評価者ブラインドは研究の強みである。 16



17

結果

主要定量結果

アウトカム	全体	ジュニア <7年	シニア ≥7年	解釈
10日ドラフティング	+0.34 SD, p=.03	+0.58 SD, p=.04	+0.22 SD, p=.24	AI利用時の品質向上
90日ドラフティング	+0.38 SD, p=.01	+0.60 SD, p=.02	+0.30 SD, p=.08	効果は持続
10日 作業時間	約−10分, p≈.05	より短縮傾向	小さい短縮	対照平均約110分
90日 作業時間	約−10分, p=.27	—	—	有意でない
90日 AIなしredlining	+0.32 SD, p=.04	−0.03 SD, p=.89	+0.45 SD, p=.02	技能移転はシニアに集中
senior–junior redlining交互作用	—	—	—	p=.04

[18]

原論文の要約で95%信頼区間は主要効果について明記されていないため、ここでは**原典に報告されたp値を優先し、CIは捏造しない**。p値と効果量から正規近似でCIを逆算することは可能だが、実際の研究はfirm構造やrating数を調整した回帰を使っているため、その逆算値を「論文報告CI」と扱うべきではない。

ドラフティングのAI効果は、優秀者をさらに突出させるというより、下位のアウトプットを減らす形で現れた。10日では下位quintileの割合が約19ポイント低下し（p<.01）、90日では約11ポイント低下したが後者は有意ではなかった（p=.12）。90日ドラフティングではAI群の分散も小さくなった（Levene p=.03）。 [13]

ジュニアではドラフティングの点推定値がシニアより大きかったものの、junior–seniorの効果差は10日でp=.15、90日でp=.17であり、統計的には「若手のほうが効果が大きい」という仮説を確証していない。PsyPost記事の表現はこの点を比較的慎重に、「appeared larger」と記述している。 [16]

AIなしredliningの意味

最終redliningでは全体で+0.32 SD、シニアで+0.45 SDだった。Googleの解説では、シニアAI群は表面的な文章修正よりも請求項構造の再設計、権利範囲を不必要に狭める表現の除去、法理に基づく修正を行う傾向があったとされる。インタビューではAIを完成稿生成器ではなく「logic auditor」のように使ったという説明も報告されている。 [15]

一方、ジュニアでは平均効果は−0.03 SDで事実上ゼロだった。さらに、下位quintileが約22ポイント増加（p=.04）、中間的な第4quintileが約42ポイント減少（p=.01）、比較的良好な第2quintileが約19ポイント増加したものの、この増加はp=.20で有意でなかった。分散増大はLevene p=.06だった。これは「全員が悪化」ではなく、成果が二極化したことを意味する。 [13]

統計解析と妥当性評価

分析モデル

著者の解説によれば、主要推定は法律事務所間の差を調整し、各弁護士が受けた評価件数の微小な違いも補正している。アウトカムは対照群を基準に標準化され、治療効果は標準偏差単位で報告された。 [19]

ただし、公開検索で確認できた原典情報だけでは、回帰式の全変数、標準誤差のクラスタリング単位、firm固定効果とstratum固定効果の細部を完全には再現できない。そのため、ここは **「firm差とrating数を調整した回帰」までを確認済み、詳細仕様は未記載として扱う** のが安全である。 [19]

欠測・脱落

133人がランダム化されたが、全課題を終えたのは91人である。したがって最終アウトカムのcomplete-case率は約68%であり、RCTとして無視できない脱落である。PsyPostおよび研究要約では、明確な処置群間の差別的脱落が主要結論を説明したとの証拠は示されていないが、attritionそのものは精度と一般化可能性を低下させる。 [20]

欠測値を多重代入した、逆確率重み付けした、あるいはLee boundsを行ったという記述は、公開検索で確認できた資料では見つからなかった。したがって、本レポートでは **欠測データの高度な補正は「確認できず」** とする。

サンプルサイズ計算

AEA登録には研究設計とアウトカムが登録されているが、公開検索で取得できた範囲では、想定効果量、α、power、ICC等を用いた事前サンプルサイズ計算の詳細は確認できなかった。よって **「事前power calculationの詳細は未確認」** とする。 [7]

これは特にjunior subgroupに重要である。最終redlining完遂者91人のうちジュニアは24人しかおらず、「平均効果なし」から「真の効果が完全にゼロ」と結論するには検出力が不足している可能性が高い。 [21]

感度分析

研究は比較的有用な複数のrobustness checkを行っている。経験年数のカットオフを3〜11年に変えても、シニアのredlining効果は概ね0.38〜0.46 SDで5%水準以下、ジュニアは有意でなかった。2,000回のランダム割付 permutation test でも、全体p=.05、シニアp=.02、ジュニアp=.93だった。 [13]

最大の感度問題はAI不使用違反である。91件中15件がAI利用の疑いありと監査でflagされた。flagを共変量として調整するとシニア効果0.44 SD、p=.02だったが、15件を除外するとシニアは0.39 SD、p=.06となった。全体効果は0.25 SD、p=.14まで弱まった。したがって「シニア効果の符号と大きさは比較的安定、統計的有意性は境界的」という評価が妥当である。 [8]

統計的結論の評価

この研究は、**AI利用時の品質向上**については比較的強い因果推論を許す。ランダム化、同一課題、blind評価、複数時点で結果が一貫しているためである。 [22]

一方、**「技能獲得」**という解釈は一段弱い。治療前のAIなし技能テストがないため、「90日後にAI群のほうが優れている」ことはランダム化によって因果推論できても、「個々人の技能がどれだけ増えたか」というchange scoreを直接観測していない。またredliningはドラフティングとは別タスクであり、transferを測っている点は長所だが、「一般的な特許専門性」全体の指標ではない。 [13]

議論

内的妥当性

強みは、事前登録、firm内ランダム化、現役専門家、3か月の実務利用、AIありとAIなしを分離した測定、独立blind評価者である。単発のオンライン実験よりも、現実の専門技能形成を検証するデザインとしてかなり優れている。 [23]

主要な脅威は五つある。第一にattrition、第二にAI不使用の遵守を直接監視できないこと、第三にbaseline skill testがないこと、第四に介入ツールが期間中に更新されたこと、第五に企業スポンサーと著者所属の利益相反である。Googleは研究費を負担しツールを開発し、7著者中6人がGoogleの従業員または関係者である。結果の不正を意味するものではないが、独立再現性を強く求める理由になる。 [4]

外的妥当性

参加者は一般的な弁護士集団ではなく、Googleのような高度な技術企業の特許案件を扱う11の米国IP事務所である。特許明細書は、技術理解、法的戦略、用語選択、権利範囲の設計を同時に要する特殊な高技能業務であり、契約レビュー、訴訟、M&A、一般企業法務へそのまま一般化できない。 [24]

さらに介入ツールは汎用ChatGPTではなく特許向けに調整されたInFlowである。モデル能力は急速に変化するため、著者自身も本結果を2024〜2025年頃のAI能力のスナップショットと位置づけている。 12

「スキル形成」と「代替」の分岐点

関連証拠を統合すると、AIが**技能形成**として働く条件は、「利用者が既にエラーを識別できるmental modelを持ち、AI出力を評価・反証・修正する」ことにある可能性が高い。逆にAIが**代替**として働く条件は、「完成稿を短時間で得ることが主目的になり、問題分解・試行錯誤・誤り訂正をAIに外注する」場合である。これは本研究のシニア対ジュニア差と整合的だが、因果機序として直接実証されたわけではない。 25

特にAnthropicの2026年RCTでは、新しいPythonライブラリを学ぶソフトウェア開発者52人にAIを使わせると、後続の理解度テストが対照群より約17%低かった。一方、AIを説明や概念理解のために利用した参加者では習得の低下が相対的に小さかった。これは「AIの有無」より「AIとの認知的関わり方」が重要だという仮説を補強する。 26

特許実務への含意

第一に、AIは「平均品質の底上げ」に強い可能性がある。本RCTでは品質向上は主にpoor workの減少として現れ、exceptional workの増加は限定的だった。したがって特許事務所では、初稿生成・表現整備・整合性チェック・従属項展開などにAIを使う合理性が高い。 13

第二に、若手の研修設計を従来どおりにすると、短期生産性と長期技能形成が衝突する可能性がある。「白紙から書く」「誤ったクレームを自分で直す」「なぜその限定を入れるか説明する」といった、非効率だが学習価値の高い仕事までAIに任せると、シニアになるまでに必要な判断材料が蓄積されない恐れがある。これは現時点ではリスク仮説だが、BastaniらのRCTとAnthropicのプログラミングRCTは同方向の証拠を示す。 27

第三に、AIを禁止するのではなく「AIを使わない評価」を定期的に残すべきである。本研究の最大の示唆は、AI使用時のアウトプットだけでは本人の技能を測れないという点にある。企業の人事評価・研修評価でも、assisted performanceとunassisted competenceを分離する必要がある。 15

倫理・法的懸念

日本では、日本弁理士会が2025年に「弁理士業務AI利活用ガイドライン」を公開している。同ガイドラインは、生成AI出力の正確性は保証されず、弁理士本人が確認し最終責任を負うべきであること、AI出力を精査せず依頼者へ提供すれば善管注意義務上の問題となり得ること、機密情報・非公開情報・個人情報の入力には慎重な管理が必要であることを明示している。 28

また日本弁理士会は、AIサービスが特許性判断や出願書類作成等を行う場合も、行為の内容によっては弁理士法75条との関係が生じると説明している。したがって「AIだから専門資格規制の外」という整理はできない。 29

実務上は、未公開発明を外部LLMへ入力することによる営業秘密・NDA・国外移転・データ再利用のリスク、誤った引用・技術記載によるenablementやsupportの問題、AIが生成した発明に対する自然人発明者の認定などが重要になる。日本特許庁も、AIそのものは現行法上発明者になれず、自然人の創作的関与をどう認定するかを継続的検討課題としている。 30

関連研究レビュー

本研究の結果は、生成AI研究で繰り返し観察される「短期の生産性向上は大きいが、技能形成は別問題」というパターンに位置づけられる。

研究	デザイン・N	分野	AI	主結果	本研究との関係
Autor et al. 2026	RCT, 133	特許弁護士	InFlow	Draft +0.34〜0.38 SD、AIなしskillはシニアのみ+0.45 SD	中核研究
Noy & Zhang 2023, Science	RCT, 453	専門職文章	ChatGPT	時間−40%、品質+18%	即時生産性の強い改善
Dell'Acqua et al.	prereg. field experiment, 758	BCGコンサルタント	GPT-4	frontier内で品質・速度改善、frontier外で正答率悪化	能力境界・過信問題
Brynjolfsson et al.	staggered rollout, 5,179	カスタマーサポート	GenAI assistant	生産性+14%、初心者+34%	noviceほど即時効果大
Bastani et al. 2025, PNAS	field RCT, 約1,000	高校数学	GPT-4 tutor	利用中改善、無防備AI群はAI除去後−17%	cognitive offloading
Shen & Tamkin 2026	RCT, 52	ソフトウェア開発	AI coding	習得テスト約−17%	skill formationの負効果
Choi & Schwarcz 2025	実験	法学教育	GPT-4	下位学生で改善、上位学生では一部悪化	法分野のheterogeneity
Chen & Bao 2026	RCT, 164	法学生	LLM	無訓練利用は一部悪化、訓練で改善	AI literacyの重要性
Cui et al. 2025	3 field RCTs, 4,867	ソフトウェア開発	coding assistant	完了タスク+26.08%、若手で大	productivity equalization
Mehmood et al. 2026	field experiment, 1,559 judges	司法	JudgeGPT	targeted trainingで採用・継続利用増	導入設計の重要性

Noy & ZhangのScience論文は453人の大卒専門職を無作為化し、ChatGPT利用群で所要時間が40%短縮し品質が18%上昇した。これは本RCTの「AIあり成果物改善」と整合するが、長期技能は測っていない。 31

Dell'AcquaらのBCG実験では758人の知識労働者をGPT-4なし、GPT-4あり、GPT-4＋prompt overviewへ割り付けた。AI能力範囲内の18課題ではAI群がより多く、より速く、高品質に仕事を完了したが、AIの能力範囲外に設計した課題ではAI利用者の正答率が約19%低下した。これは「専門家の判断がAIの誤りを上書きできるか」という本研究の論点と直結する。 32

Brynjolfssonらの5,179人のカスタマーサポート研究では、AI導入により平均生産性が約14%、初心者・低技能者では約34%改善した。これはAIが即時成果を平準化するという強い観察証拠である。ただしRCTではなく段階導入デザインであり、本研究より技能形成に関する因果識別は弱い。 [33]

Bastaniらの約1,000人の高校生フィールド実験では、通常型GPT-4を使った練習中の成績は48%改善した一方、AIを外した試験では対照群より17%低下した。学習を促すガードレール付きTutorではこの悪影響が大幅に抑制された。これは「AIの設計が代替にも教育にもなりうる」ことを直接示す重要な比較対象である。 [34]

ソフトウェア分野では、Microsoft、Accenture等の4,867人を含む3件のfield RCTを統合した研究で、AI coding assistantにより完了タスク数が26.08%増え、経験の浅い開発者ほど効果が大きかった。一方、別の小規模RCTではAI利用者の後続理解度が約17%低かった。短期生産性と長期能力が逆方向に動き得るという本研究の中心命題を支持する組み合わせである。 [35]

法分野では、Choi & SchwarczのGPT-4実験で、単純な多肢選択では成績が改善した一方、複雑なessayでは明確な改善がなく、初期技能の低い学生ほど恩恵が大きかった。Chen & Baoの164人RCTでは、LLMアクセスのみの群は答案が短くなり、判例の誤記が増え、訓練を付けた群のほうが有効利用しやすかった。 [36]

2026年のパキスタン司法研究では1,559人の裁判官を対象にJudgeGPTのfield experimentを行い、AIそのものだけでなく**targeted training**が採用・利用量・継続使用に重要だった。これは職場導入における「ライセンス配布」だけでは不十分であることを示す。 [37]

追加解析案

元データが利用可能であれば、まず「技能形成」をより直接に評価できる解析へ組み直す価値がある。必要なデータは、個人ID、firm、経験年数、職位、事前AI使用歴、10日・90日の各ratingの生データ、各評価者ID、作業時間、InFlowの利用ログ、prompt履歴、編集履歴、脱落理由、redliningのAI利用監査結果である。

最優先は、ratingを単純平均せず、**lawyer・rater・firmを階層化したmixed-effects model**を使うことである。たとえば、アウトカムをquality score、固定効果をtreatment、experience、treatment×experience、時点、事前AI経験とし、lawyer・firm・raterにrandom interceptを置けば、rating数の不均衡をより自然に処理できる。

第二に、経験7年という二分法だけでなく、**経験年数を連続変数としてrestricted cubic spline**でモデル化し、AIが技能形成にプラスへ転じる「経験年数の転換点」が存在するか推定すべきである。3〜11年の閾値感度分析は有用だが、連続情報を切断するより統計効率が高い。

第三に、attritionについて、群別脱落率と脱落理由を公開し、**inverse probability weighting、multiple imputation、Lee bounds**などを併用するべきである。特に「忙しい優秀な弁護士ほど最終課題を辞退した」などの選択があるとcomplete-case推定は歪む。

第四に、AI利用ログを用いて、単なるITTではなく、**利用量・利用様式を媒介変数とするcausal mediation analysis**を行うべきである。「AIに全文を書かせる」「説明を求める」「批判させる」「自分の稿を検証させる」といった使用モードがskill transferを予測する可能性がある。ただし利用量はランダム化されていないため、単純なusage-performance相関は因果と解釈すべきでない。

第五に、redliningの得点分布がジュニアで二極化したため、平均値だけでなく **quantile treatment effects、distributional regression、variance treatment effect** を事前登録して検証するべきである。ジュニアの「平均0」は、実際には「一部が大きく伸び、一部が悪化」の混合である可能性がある。

第六に、最も重要なのは次回RCTで**治療前のAIなしredlining・draftingテスト**を取得することである。これによりANCOVAまたはdifference-in-differences的な設計で、単なるendline群差ではなく個人単位のskill gainを推定できる。

結論・提言

本RCTから言える最も確かなことは、生成AIが特許ドラフティングの成果物品質を約0.3〜0.4標準偏差改善し、とりわけ低品質アウトプットを減らすという点である。これ自体は実務的に意味のある効果である。[9]

しかし、AI利用時の生産性向上を「人間が上達した」と解釈するのは誤りである。AIを外した90日後の専門判断課題で改善が明確だったのはシニアであり、ジュニアには平均改善がなかった。ただしジュニアが劣化したと確定したわけでもなく、サンプルが小さいため二極化の原因は未解明である。[13]

日本の特許・知財実務への含意は、AI導入の是非よりも**研修設計**にある。日本弁理士会自身も生成AIを生産性向上に使える一方、正確性確認、秘密保持、個人情報、最終的な専門家責任を求めている。[28]

政策・企業向けの提言は次のとおりである。

- **新人教育ではAIなし演習を残す。** 初稿作成、claim construction、redliningを一定割合で手作業評価し、assisted outputとunassisted competenceを分離する。
- **AIを「生成器」より「批判者」として使う研修を増やす。** 根拠説明、反証、誤り探索、代替claim作成をAIに要求する形式のほうが技能形成と両立する可能性が高い。
- **人事KPIを時間短縮だけにしない。** 短期生産性だけを報酬化すると、若手が学習努力を外部化するインセンティブを生む可能性がある。
- **機密保持・ログ管理・モデル更新管理を必須化する。** 未公開発明、顧客データ、NDA対象情報を扱うため、企業向け閉域環境と監査証跡を標準とすべきである。
- **次世代の研究ではbaseline skill testと1〜3年追跡を必須にする。** 3か月では専門家形成について結論するには短く、昇進・独立判断・AI監督能力まで測る必要がある。
- **独立再現研究を行う。** Google製ツール、Google取引事務所、Google主導研究という構造上、別ベンダー・別顧客・別法域での再現性確認が不可欠である。[24]

---

[1] [9] [10] [22] Does AI Assistance Enhance or Erode Expertise? Evidence from a Three-Month Field Experiment in Patent Drafting
https://www.nber.org/papers/w35720?trk=article-ssr-frontend-pulse_x-social-details_comments-action_comment-text&utm_source=chatgpt.com

[2] Does AI Assistance Enhance or Erode Expertise? Evidence from a Three-Month Field Experiment in Patent Drafting
https://research.google/pubs/does-ai-assistance-enhance-or-erode-expertise-evidence-from-a-three-month-field-experiment-in-patent-drafting/?utm_source=chatgpt.com

[3] [8] [13] [16] [18] Does AI Assistance Enhance or Erode Expertise? Evidence from a Three-Month Field Experiment in Patent Drafting – Read with an AI tutor
https://www.chapterpal.com/s/z2mfnoz7/does-ai-assistance-enhance-or-erode-expertise-evidence-from-a-three-month-field-experiment-in-patent-drafting?utm_source=chatgpt.com

4 5 14 20 24 Prior expertise may determine whether AI builds skills or substitutes for them, field experiment suggests
https://www.psypost.org/prior-expertise-may-determine-whether-ai-builds-skills-or-substitutes-for-them-field-experiment-suggests/?utm_source=chatgpt.com

6 〖海外記事紹介〗AIの創造性には数学的な上限が存在する ― 人間のアマチュアレベルを超えることはできない
https://techfeed.io/entries/69261a603837533862ef3ba7?utm_source=chatgpt.com

7 17 23 AEA RCT Registry
https://doi.org/10.1257/rct.15823-1.0?utm_source=chatgpt.com

11 15 25 Does better work always mean better workers?
https://research.google/blog/does-better-work-always-mean-better-workers/?trk=article-ssr-frontend-pulse_little-text-block&utm_source=chatgpt.com

12 Better work, not always better workers | CEPR
https://cepr.org/voxeu/columns/better-work-not-always-better-workers?trk=public_post_comment-text&utm_source=chatgpt.com

19 Better work, not always better workers | CEPR
https://cepr.org/index.php/voxeu/columns/better-work-not-always-better-workers?utm_source=chatgpt.com

21 Research Roundup - September 2026 - Publications - The Technology Policy Institute
https://techpolicyinstitute.org/publications/artificial-intelligence/research-roundup-september-2026/?utm_source=chatgpt.com

26 How AI assistance impacts the formation of coding skills \ Anthropic
https://www.anthropic.com/research/AI-assistance-coding-skills?_bhlid=b2c4508f2c65421e278407de6d382d3d8caa468b&utm_source=chatgpt.com

27 34 Generative AI Without Guardrails Can Harm Learning: Evidence from High School Mathematics
https://doi.org/10.2139/ssrn.4895486?utm_source=chatgpt.com

28 弁理士業務 AI 利活用ガイドライン
https://www.jpaa.or.jp/cms/wp-content/uploads/2025/04/AIservices-guideline.pdf?utm_source=chatgpt.com

29 資格を持たずに弁理士業務を行う者（非弁理士）にご注意ください | 日本弁理士会
https://www.jpaa.or.jp/intellectual-property/protect/non_patent_attorney/?utm_source=chatgpt.com

30 資料１ 審査基準に関する最近の取組について 令和７年１１月１７日
https://www.jpo.go.jp/resources/shingikai/sangyo-kouzou/shousai/kijun_wg/document/18-shiryou/003.pdf?utm_source=chatgpt.com

31 Experimental evidence on the productivity effects of generative artificial intelligence
https://doi.org/10.1126/science.adh2586?utm_source=chatgpt.com

32 Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of Artificial Intelligence on Knowledge Worker Productivity and Quality
https://papers.ssrn.com/sol3/Delivery.cfm/4573321.pdf?abstractid=4573321&mirid=1&utm_source=chatgpt.com

33 Generative AI at Work
https://www.nber.org/papers/w31161?utm_source=chatgpt.com

35 The Effects of Generative AI On High-Skilled Work:
https://economics.mit.edu/sites/default/files/inline-files/draft_copilot_experiments.pdf?utm_source=chatgpt.com

36 AI Assistance in Legal Analysis: An Empirical Study
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4539836.&utm_source=chatgpt.com

37 Courts of Tomorrow: Evidence from a Nationwide Rollout of Generative AI
https://papers.ssrn.com/sol3/Delivery.cfm/7119447.pdf?abstractid=7119447&mirid=1&utm_source=chatgpt.com