AI はベテランを鍛え、若手を代替するのか

特許弁護士 133 人・3 か月 RCT（Autor ほか、NBER WP 35720）の一次情報検証と知財実務への示唆

2026 年 11 月 8 日

凡例　本書では、一次情報で確認した事実と、分析・推論を区別して記載する。見出しまたは段落冒頭に【分析】と付した部分は推論・提言であり、それ以外は出典に基づく事実記載である。出典は本文中に上付き番号［n］で示し、詳細を文末の参考文献に記した。各文献には確認状況（本文確認／補助調査に基づく／未確認 等）を付した。

要旨

- 元論文は、David Autor（MIT）と Google 所属の共著者 6 名による NBER ワーキングペーパー35720「Does AI Assistance Enhance or Erode Expertise?」（2026 年 9 月発行、10 月改訂）であり、**査読前**である[1,2]。米国の知財系法律事務所 11 所・133 人を対象とした、事前登録済みの 3 か月ランダム化比較試験（RCT）である点は記事どおり[1,3]。
- AI ありの明細書ドラフト作成では、品質が 10 日目に+0.34SD、90 日目に+0.38SD 向上し、伸びは若手のほうが大きく見えた[1]。
- 一方、90 日目に AI なしで行った赤入れ（レッドライン）課題で差が出たのは**ベテラン（経験 7 年以上）だけ**で+0.45SD。若手は平均で効果がなく（−0.03SD）、点数が「悪い」と「良い」の両方向に**二極化**した[1,2]。
- 記事見出しの「若手の思考を代替しうる」は、論文の「一部の若手では代替した可能性」という解釈を一般化した単純化である。若手の平均的な劣化は示されていない[2]。
- TechFeed 本文は PsyPost 記事の要約として全体的に正確である。ただし、「シニアは品質、ジュニアは速度」とするメカニズムの説明は論文の推定値と合わない。また、AI 使用疑いを除外した後の「シニア優位の維持」もやや楽観的である（除外後 p=0.06）[2,6]。
- 【分析】AI 導入の効果は「AI ありの成果物の質」と「AI なしの判断力」に分けて測るべきである。若手には「自分で起案してから AI に批評させる」順序を徹底し、AI なしの定期チェックで下位層の増加を監視するのが妥当。ただし、小標本、Google 資金・Google 製ツー

ル、主要な異質性分析が事前登録外である点から、「若手に AI を使わせるな」という結論の根拠にはならない。

Autor はこの結果を、AI は「成果の平準化装置（performance equalizer）」であると同時に「スキルの格差拡大装置（skill-disequalizer）」でもある、と表現している（報道ベース）[7]。方向性としては興味深い知見だが、AI なし課題の完了者が 91 人にとどまること、ベテラン効果の一部の頑健性検定で p 値が 0.06〜0.09 まで弱まることから、**暫定的なシグナル**として扱うべきである[2]。

1. 元論文の特定

表 1　元論文の書誌情報

項目	内容
タイトル	Does AI Assistance Enhance or Erode Expertise? Evidence from a Three-Month Field Experiment in Patent Drafting[1]
著者・所属	David Autor（MIT 経済学部・NBER、Google Technology & Society Visiting Fellow）、Tanya Rodchenko、Josh Martin、Zanna Iscenko、Scott Strand、David Pearl、Melissa Ferere（いずれも Google LLC）[1,4]
掲載先	NBER Working Paper No. 35720（査読前のワーキングペーパー）。SSRN にも掲載[1]
公開日	2026 年 9 月発行（初版 9 月 3 日付）、2026 年 10 月改訂。MIT 経済学部サイト掲載の PDF は 2026 年 10 月 7 日付[1,2]
DOI	10.3386/w35720
事前登録	AEA RCT Registry AEARCTR-0015823（2025 年 5 月 21 日公開、登録題名 "AI & Expertise research with patent lawyers"）[1,3]
資金・利益相反	実験の直接費用は Google が負担し、実施主体も Google。MIT の倫理委員会は「MIT は本研究に関与していない」と判断し、MIT 側の審査は行われていない。NBER の開示によれば、Rodchenko・Iscenko・Pearl・Ferere は Google 正社員（株式報酬あり）、Strand は正社員かつ株主、Martin は Google と有償の契約関係にある[1,2]
付随発表	Google Research Blog「Does better work always mean better workers?」（2026 年 10 月 7 日、Autor・Rodchenko 連名）[4]

PsyPost の記述の照合：PsyPost（Eric W. Dolan、2026 年 11 月 7 日）[5]は、査読前であること、Google が実施・資金提供したこと、主な限界（小標本、事前テストなし、AI 不使用を検証できないこと、3 か月という短さ）をおおむね正確に伝えている。ただし「著者 7 人中 6 人が Google で働く」との記述は、厳密には正社員 5 人と有償契約者 1 人（Martin）である[1]。

2. 記事の記述と元論文の照合

TechFeed 記事[6]（TechFeed 公式）は、PsyPost の記事[5]（11 月 7 日）を紹介する日本語の要約記事である本文テキストで全文を確認した。2.1 では見出しの各要素を、2.2 では本文の主な記述を元論文と照合する。

2.1 見出しの照合

表 2　見出しの記述と元論文の照合結果

見出しの記述	元論文の内容	判定
特許弁護士 133 人	推薦された 156 人のうち、オンボーディングを終えてランダム化されたのが 133 人（介入群 90 人、対照群 43 人）。10 日目課題の完了者 105 人、90 日目ドラフト 98 人、**AI なし赤入れ 91 人**[2]	正確。ただし核心の結果（赤入れ）は 91 人に基づく
3 か月	各コホートに約 90 日間のアクセスを付与。全コホートの実施期間は 2025 年 5 月～2026 年 2 月[2]	正確
ランダム化比較試験	事前登録あり。事務所内で 2:1 に割り付け、事務所と経験年数（事務所内の中央値）で層化[2,3]	正確。ただし若手・ベテラン別の分析は事前登録の主要アウトカムに含まれない（3.2 参照）
ベテランの判断力を鍛え	AI なし赤入れでベテランは+0.45SD（p=0.02。並べ替え検定でも p=0.02）[1,2]	概ね妥当。ただし「鍛えた」は推論（事前のスキル測定がなく、無作為化による群間バランスが根拠）。影響の大きい 1 事務所を除くと 0.33SD（p=0.08）、AI 使用疑いの提出物を除くと 0.39SD（p=0.06）に弱まる[2]
若手の思考を代替しうる	若手は平均－0.03SD（p=0.89）で効果なし。点数は二極化し、「悪い」+22pt、「凡庸」－42pt、「良い」+19pt（非有意）。論文は「一部の若手では習得	**単純化**。「しうる」の留保は論文の解釈の範囲内だが、平均的な劣化は示されておらず、「良

見出しの記述	元論文の内容	判定
	を加速し、別の若手では代替した可能性と整合的」とし、理由は答えられないと明記[2]	い」点数の増加という半面が落ちている

【分析】PsyPost の原題は、事前の専門性によって AI がスキルを「築く（builds）」か「代替する（substitutes）」かが決まりうる、という一般的な表現である。TechFeed の見出しは、これを「ベテランを鍛え／若手の思考を代替」と年次の対比に当てはめており、原題より踏み込んでいる。

2.2 本文の照合

表 3　TechFeed 本文の主な記述と元論文の照合結果

TechFeed 本文の記述	元論文の内容	判定
24 事務所に打診し 11 事務所が参加。90 人を AI 群、43 人を対照群に割り付け、3 か月追跡	一致[2]	正確
Google Labs 開発の未公開ツール（プロンプトによる修正、複数視点のフィードバック、矛盾チェック）。対照群は生成 AI なし	一致[2]	正確
10 日目ドラフトで品質が「約 0.33 標準偏差」向上、時間は約 10 分短縮（対照群平均 110 分）。改善の中心は弱いドラフトの減少	品質は+0.34SD（p=0.03）[1]。時間と分布に関する記述は一致[2]	概ね正確（0.33 は論文要旨の 0.34 と小差。どこから来た数値かは未確認）
90 日目の AI なし赤入れで、シニアは明確に上回り、ジュニアは平均差なし。ジュニアの分布は「平凡」が減り「低い」「良い」が増加	ベテラン+0.45SD（p=0.02）、若手−0.03SD。「凡庸」減、「悪い」「良い」増[1,2]	正確
AI 使用中に最も恩恵を受けたジュニアが、AI なしでは優位を示さなかった。本人は実力向上を主観的に感じていた	論文要旨は、AI の最大の恩恵が最も定着しなかった層に生じたとする[1]。若手の満足度+1.48SD、主観的品質+0.95SD[2]	正確

TechFeed 本文の記述	元論文の内容	判定
著者らの結論として "Foundational expertise may be a prerequisite for extracting durable skill from AI-assisted practice" を引用	論文要旨の末文と一致[1]	正確
メカニズム：シニアは AI 使用中に速度より品質の向上が目立ち、ジュニアは速度が上がった。シニアは AI を「考えるための道具」とし、ジュニアは AI 出力をそのまま受け取って試行錯誤を省略した可能性	AI ありのドラフトで品質・速度とも伸びが大きかったのは若手（品質+0.58／+0.60SD、時間約 18 分短縮）。ベテランのドラフト品質向上（+0.22／+0.30SD）は有意でない[2]。論文は理由に答えられないとし、Google ブログの定性分析では若手の弱点は対照群でも同じだった[2,4]	**要注意**。「シニア＝品質、ジュニア＝速度」という対比は主要推定値と合わない。「出力をそのまま受け取った」は記事側の推論（「可能性」の留保はある）
AI 使用疑いは 91 件中 15 件。除外後もシニア層の優位は維持され、全体の差は有意性が消えた	除外後、全体 0.25SD（p=0.14）、ベテラン 0.39SD（p=0.06）[2]	概ね正確。ただしベテランも p=0.06 で 5%水準では有意でなく、「維持」はやや強い表現
著者 7 人中 6 人が Google 在籍	正社員 5 人と有償契約者 1 人（Martin）[1]	厳密には不正確（PsyPost と同じ記述）
限界として、特許弁護士のみ、小標本（特にジュニア）、事前テストなし、3 か月の短さ、利益相反、査読未了を列挙	著者自身が挙げる限界および開示内容と一致[1,2]	正確

【分析】TechFeed 本文は、PsyPost 記事の要約として全体的に正確であり、限界や利益相反にも触れている。問題は 3 点ある。(1) 見出しが原題より踏み込んでいる。(2) メカニズムの説明で、「シニアは品質、ジュニアは速度」という論文の推定値と合わない対比を示している（PsyPost 側の記述に由来するかは未確認）。(3) AI 使用疑いを除外した後の頑健性を、やや楽観的に伝えている。

3. 研究デザイン

3.1 対象・介入・評価

- **対象**：Google と継続的に（非独占で）特許起案の取引がある米国の知財系法律事務所 24 所に打診し、11 所が参加。1 事務所の 2 コホートを別扱いとして計 12 コホート（1 コホート 3〜49 人）。平均経験年数 12.4 年、7 年以上の「ベテラン」が 89 人（67%）。参加者には契約上の時間単価で報酬が支払われた[2]。
- **若手・ベテランの区分**：割り付けの層化には事務所内の経験年数の中央値を使い、分析では「7 年以上をベテラン」とした（Google 社内弁護士の助言による業界慣行）。閾値を 3〜11 年のどこに置いても、ベテランの赤入れ効果は 0.38〜0.46SD で 5%水準で有意[2]。
- **介入**：Google Labs の未公開の文章作成支援ツール。Patently-O は名称を「InFlow」と報じ[8]、Google のブログは現在「Gemini Notebook」の一部になったと説明している[4]。主な機能は、プロンプトによる文章修正、複数の AI 編集者ペルソナによるフィードバック、文書内の不整合のルールベース検査で、2025 年 7 月に赤入れ機能が追加された。AI 基礎研修は両群が受け、ツール専用研修（特許向けプロンプト集を含む）は介入群のみが受けた。したがって介入は「アクセス＋研修」の複合である[2]。
- **対照群**：期間中、生成 AI の文章ツールは提供されず、通常の法務ツールのみを使用[2]。
- **課題**：(1) 10 日目ドラフト：独立クレームと図面・発明者メモをもとに従属クレーム 5〜7 項と詳細な説明を作成（2 時間以内・1,000 語以内。発明はモデル学習に使っていない hold-out データから選定）。(2) 90 日目ドラフト：より難しい課題で 2,000 語以内。(3) 90 日目赤入れ：誤りを含む AI 生成の明細書草稿を **AI なしで**添削。先行技術調査や拒絶理由対応（中間処理）は対象外[2]。
- **評価**：独立系知財事務所に所属する評価者 6 人のうち 2 人を無作為に割り当て、群を知らせないブラインド状態で採点。ルーブリックは Google 社内弁護士と共同で作成し、権利行使可能性、技術的正確性、戦略的曖昧性（USPTO の最広合理的解釈を踏まえた広いクレーム文言）、網羅性、明瞭性の 5 次元[2,4]。評価者 2 人の相関は 0.59／0.47／0.55（中程度）。補助として Gemini 2.5 Flash による LLM 採点も行ったが、専門家との相関は 0.47／0.19／0.14 と低い[2]。

- **作業時間**：守秘の制約から自己申告（ストップウォッチ記録と事後アンケート）。**事前のスキル測定**は、著者は望んだが参加事務所が拒否したため実施されていない[2]。

3.2 事前登録との齟齬

- 登録された主要アウトカムは、品質（第三者評価）、作業時間、主観的な品質・時間の変化、満足度。経験年数による異質性分析とその閾値は**事前に特定されていない**[3]。
- 登録時の設計では、赤入れは群ごとに AI あり／なしで行う想定だった。「全員が AI なしで赤入れ」する課題は後から設計された[3]。
- 計画サンプルは「最大 350 人」で、実際の 133 人はその約 38%。分析計画は公開されていない（非公開情報が存在する旨の表示あり）[3]。

【分析】したがって、論文の見出しになっている若手・ベテラン間の逆転は、確認的な結果ではなく**探索的な分析**と位置づけるのが妥当である。

4. 主な結果

表 4　処置効果の推定値（標準偏差単位）[1,2]

アウトカム	全体	若手（7 年未満）	ベテラン（7 年以上）
10 日目ドラフト（AI あり）	+0.34SD（p=0.03）	+0.58SD（p=0.04）	+0.22SD（p=0.24）
90 日目ドラフト（AI あり）	+0.38SD（p=0.01）	+0.60SD（p=0.02）	+0.30SD（p=0.08）
90 日目赤入れ（AI なし）	+0.32SD（p=0.04）	−0.03SD（p=0.89）	**+0.45SD（p=0.02）**

- **ドラフトの改善の中身**：最上位の点数が増えたのではなく、下位の点数が減ったことによる改善（10 日目の最下位五分位は−19pt）。若手とベテランの効果差自体は有意でない（p=0.15、0.17）[2]。Google のブログはパーセンタイル換算で、ドラフト+10〜11pt、赤入れ全体+9pt、ベテラン+13pt としている[4]。
- **赤入れでの逆転**：効果はベテランのほうが有意に大きい（片側 p=0.04）。対照群では若手とベテランに差がなかったのに、3 か月の AI 利用後には AI なしの判断力に差が開いた、というのが論文の主張である[2]。

- **若手の二極化**：分散が拡大し（Levene 検定 p=0.06）、「悪い」点数の比率が+22pt（p=0.04）、「凡庸」が－42pt（p=0.01）、「良い」が+19pt（p=0.20）。最上位は増えていない[2]。
- **速度**：10 日目ドラフトは対照群平均 110 分に対し約 10 分短縮（p=0.05）、若手は約 18 分短縮。赤入れでは時間短縮は見られない[2]。
- **自己認識とのずれ**：介入群の若手は、90 日目の満足度が+1.48SD、10 日目の主観的品質が+0.95SD と高く自己評価したが、AI なしの成績は伸びていない[2]。Autor はこれを「有能さの錯覚（illusion of competence）」と呼んでいる[7]。
- **メカニズム**：論文は、効果が人によって正負に分かれる理由には「答えられない」と明言している[2]。Google のブログの定性分析によれば、若手は両群とも冒頭から順に重要度の低い文章を直し、同義語の置換に終始し、欠陥を見つけてもコメントで指摘するだけで修正しない（diagnose-without-execute）傾向があった。これは対照群でも同じで、若手の基礎的な弱点が AI 利用で改善されなかったことを示す。ベテランは AI を「logic auditor（論理監査役）」として使ったと語り、赤入れでは重要度の低い文章を飛ばしてクレームを組み直し、修正に法理を紐づけていた[4]。論文は、時間と努力の投入が学習を左右するという仮説（中国の中高生を対象とした Strömberg ほかの研究）と整合的だとしつつ、直接の検証はしていない[2]。

5. 限界と批判的検討

5.1 著者自身が認める限界

小標本（とくに若手）、Google の取引先という洗練された事務所に偏ったサンプル、事前スキル測定の欠如、3 か月という短い観察期間、赤入れでの AI 不使用を直接検証できないこと[2]。Google のブログはさらに、モデルの進歩や AI への慣れが進んだ現在に再実施すれば結果が変わりうることも認めている[4]。

5.2 【分析】追加の検討

(1) **検出力と効果の誇張リスク**：若手で AI なし赤入れを完了したのは 20 人台にすぎない（二次解説では 24 人、うち対照群 7 人[13]）。独立した検討（野崎秀吾氏の note 記事）は、0.45SD に対するサブグループの検出力を約 34〜47%と見積もり、Type M 誤差（有意にな

った推定値が真の効果を過大に示す誤差）のリスクを指摘している[12]。ベテランの+0.45SDも、真の効果より大きく出ている可能性がある。

(2) **頑健性の濃淡**：AI 使用が疑われる 15 件を除くと、全体効果は 0.25SD（p=0.14）で有意でなくなり、ベテランは 0.39SD（p=0.06）となる。疑いのある提出物は対照群でも見つかっており（29 件中 4 件）、AI 不使用ルールの遵守は群を問わず不完全だった可能性がある[2]。

(3) **測定の妥当性**：評価者間相関は中程度で、LLM と専門家の相関は 90 日課題で 0.14〜0.19 と低い。赤入れは単一の模擬課題であり、実際の出願の品質（登録率や係争での生存）とは結びついていない[2]。

(4) **外的妥当性**：米国実務（USPTO の最広合理的解釈）、英語、Google という単一の高度な顧客という条件下の結果である。日本の弁理士実務（中間処理の比重、日本語明細書、JPO の審査基準）にそのまま当てはまる保証はない。

(5) **利益相反**：実施主体・資金・ツールのいずれも Google である[1]。ただし結果は自社ツールに一方的に有利な内容ではなく（若手の学習効果なし）、事前登録もされている。Google は本研究を「Autor のフェローシップの一環としての独立研究」と説明している[4]。

6. 反響・評価

- **Autor（著者）**：Fortune への書面回答で、AI は自動的にスキルを平準化するという通念に挑む結果だと述べた。事務所が育成機会を自動化で消せば、将来のシニアパートナーを生む徒弟制のパイプラインを断ち切ると警告する一方、自らの提言は確立した解決策ではなくアイデアだと留保している[7]。
- **Tyler Cowen（Marginal Revolution）**：RCT は多くの条件を固定する必要があるため、人と業務の配置が長期的に調整されることによる利益を見落としうると指摘した[9]。
- **Dennis Crouch（Patently-O）**：AI が特許業務で重要な役割を持つことは明らかとしたうえで、「スキルを築きながら AI をどう使うか」が残る問いだとした[8]。
- **RohitAI（技術ブログ）**：若手の−0.03SD は統計的に決着しておらず、スキル喪失の証拠ではないと指摘。ドラフトと赤入れの効果を割り算して「保持率」を出すことはできないとも注意している[10]。

- **メディア**：Fortune[7]、The Decoder[11]、PsyPost[5]などが報道した。The Decoder の見出し（若手の職業的成長を「静かに阻害するかもしれない」）は「may」付きだが、平均効果ゼロという結果に対してはやや煽り気味である。日本では TechFeed 公式が PsyPost 記事の要約として紹介している[6]。

7. 先行研究との比較 ―「格差を縮める」か「広げる」か

表 5　関連研究との比較

研究	対象・規模	主な結果	本研究との関係
Brynjolfsson, Li & Raymond（QJE 2025）[14]	カスタマーサポート担当 5,179 人	生産性+14%、初心者・低スキル層は+34%	AI ありの成果での平準化と一致
Dell'Acqua ほか（BCG、2023）[15]	コンサルタント 758 人	AI の得意領域（フロンティア内）ではタスク数+12.2%、速度+25.1%、品質+40%。平均以下の層+43%、平均以上+17%。フロンティア外では正答率が 19pt 低下	平準化に加え、AI の誤りを見抜けないリスクを示す
Bastani ほか（PNAS 2025）[16]	トルコの高校 9〜11 年生 約 1,000 人（約 50 クラス、クラス単位で無作為化）	練習中は GPT Base+48%、GPT Tutor+127%。AI なしの試験では GPT Base が－17%。答えを出さない設計の Tutor では害がほぼ消失	「成果は上がるが学習は損なわれる」の先例。ツール設計で緩和可能
Otis ほか（Management Science 2026）[17]	ケニアの起業家 640 人	低業績者は AI 助言で 10%近く悪化し、高業績者は 15%超の恩恵を受けた可能性（初期の会議版では約 8%悪化）	「格差を広げる」側。本研究の赤入れ結果と同方向
Lee, Sarkar ほか（Microsoft/CMU、CHI '25）[18]	知識労働者 319 人（実例 936 件）の調査	生成 AI への信頼が高いほど批判的思考が少なく、自己への信頼が高いほど多い	「専門性が検証を可能にする」機序と整合（自己申告・相関研究）
Kosmyna ほか（MIT Media Lab、2025）[19]	54 人（第 4 セッションは 18 人）、脳波計測	LLM 群は脳の結合性が最も弱く、自分の文章を想起しにくい	方向は一致するが、査読前の小規模実験室研究
Choi & Schwarcz（J. Legal Educ.）[20]	法学生	GPT-4 の支援でクラス下位の学生は大きく向上し、上位は低下。単純な	法律分野での成果の平準化

研究	対象・規模	主な結果	本研究との関係
		択一問題では大きく向上したが、複雑な論述問題では向上せず	
Choi, Monahan & Schwarcz（Minn. L. Rev. 2024）[21]	法学生	GPT-4 は法的分析の質をわずかかつ不安定にしか改善しないが、速度は大きく一貫して向上（時間短縮 11.8～32.1%）	法律文書では速度効果が中心
Bednar ほか（2026、Northwestern U. L. Rev.掲載予定）[22]	法学生 約 100 人	初期段階での AI 利用は後の AI なし課題を必ずしも損なわない。効果は「いつ・どう使うか」で決まる	法律分野で「AI なしの事後測定」を行った先例

【分析】本研究は両説を統合する位置にある。**AI ありの成果では「縮める」説**（Brynjolfsson、Dell'Acqua、Choi & Schwarcz）を再現し、**AI なしのスキルでは「広げる」説**（Otis、Bastani）の側に入る。新しさは、「縮める／広げる」が測定対象（成果かスキルか）によって入れ替わることを、現役の専門職の実務課題で示した点にある。共通する要因は「AI出力を評価・検証する力」であり、Lee らの調査、Dell'Acqua のフロンティア外の結果、Bastani のツール設計の効果と整合する。

8. 【分析】知財実務への示唆

以下は本研究と関連研究から導いた推論であり、日本の事務所で検証されたものではない。日本弁理士会の「弁理士業務 AI 利活用ガイドライン」（2025 年 4 月）は、AI 生成物の正確性の確認と弁理士の最終責任という原則を定めるが、確認した範囲では若手育成への具体的な言及は見当たらない[23]。東京弁護士会の会報 LIBRA（2026 年 7・8 月号）や松尾剛行弁護士も、AI により定型案件という OJT の場が失われる懸念を指摘している[24,25]。

1. **成果と育成を別々に測る**：AI ツール導入の評価指標に、AI なしの定期チェックを組み込む（例：四半期ごとに、ブラインド採点で模擬明細書の赤入れ、クレーム添削、拒絶理由への応答方針の立案）。見るべきは平均点だけでなく**下位層の比率**である。本研究の若手で増えたのは「悪い」点数だった。
2. **若手には「先に自分で、後で AI」の順序を守らせる**：クレームの骨格と応答の論理はまず自力で組み立て、その後 AI を批評役（論理監査役）として使う。Bastani の研究で答えを出さ

ない設計の Tutor が害を抑えたことから[16]、社内 AI ツールも若手向けには「代わりに書く」より「指摘・質問する」設定を優先するのが妥当である。

3. **段階的に導入し、「AI なしで習得すべき業務」を明示する**：明細書の定型部分の起案や誤記チェックは AI に任せてよい一方、クレームの範囲設計、先行技術との差異の論理、中間処理での補正方針は、独立して起案できる水準に達するまで自力で行う業務と定める。一律の AI 禁止期間より、**業務ごとの線引きと到達度の確認**のほうが本研究の示唆に合う。
4. **「diagnose-without-execute」を直接鍛える**：欠陥を指摘しても直さない、重要度の低い箇所に時間を使うといった若手の基礎的な弱点は、AI では改善されなかった[4]。ベテランが赤入れ結果を「なぜそう直したか」まで添えて返す指導は、AI 時代にむしろ価値が上がる。
5. **ベテランの活用**：ベテランには AI を積極的に使わせ、論理監査役としての使い方を言語化して若手に移転する役割を担わせる。AI の誤りを題材にしたレビュー会も有効である。
6. **採用・人員計画**：AI で若手の成果物の質が上がっても、それを理由に若手採用を減らすのは短絡的である。本研究によれば、AI ありの成果は AI なしの判断力の代わりにならない。中途採用への偏重は、業界全体として将来のベテラン供給を細らせる。
7. **企業知財部門**：外部事務所の評価に「AI 使用時の品質」と「担当者の独立した判断力」の両方を入れること、社内の AI 利用ルールを経験年数・到達度に応じて段階化することが考えられる。
8. **自前の検証**：本研究は米国・英語・Google 顧客・7 年基準という条件下の結果である。日本の事務所・知財部門では、自らの業務で小規模パイロット（段階的導入、ブラインド採点、AI なし課題の事前・事後比較）を行い、自前のデータで判断するのが最善である。

9. 留保事項と情報源の確認状況

- 本研究は**査読前の NBER ワーキングペーパー**で、再現研究はまだない。若手・ベテラン別の分析は事前登録外で検出力も低いため、「ベテラン+0.45SD・若手ゼロ」という数値は今後の研究で縮小しうる[2,3]。
- 赤入れの結果は 91 人・単一課題・3 か月の観察に基づく。「若手が長期的に専門家になれない」ことは示しておらず、著者と Google のブログもその旨を明言している[2,4]。
- 使用ツールは 2025 年時点の未公開プロダクトであり、現在のモデルとは性能が異なる[4]。

- **確認状況**：NBER 論文ページ・MIT 掲載 PDF・Google ブログ・PsyPost・Fortune・Marginal Revolution・RohitAI は本文を確認。Patently-O は会員限定部分を除く冒頭のみ。AEA 登録内容と野崎氏の note 記事は補助調査に基づく。TechFeed 記事は、提供された本文テキストで全文を確認。先行研究の数値の一部（Dell'Acqua、Choi & Schwarcz、Kosmyna ほか、Lee ほか）は要旨・二次要約に基づく。

参考文献

[1] Autor, D., Rodchenko, T., Martin, J., Iscenko, Z., Strand, S., Pearl, D., & Ferere, M. (2026). "Does AI Assistance Enhance or Erode Expertise? Evidence from a Three-Month Field Experiment in Patent Drafting." NBER Working Paper No. 35720（2026 年 9 月発行、2026 年 10 月改訂）. DOI: 10.3386/w35720. 【本文確認】 https://www.nber.org/papers/w35720

[2] 同上論文 PDF（MIT 経済学部サイト掲載、2026 年 10 月 7 日版）. 【本文確認】 https://economics.mit.edu/sites/default/files/2026-10/Does%20AI%20Assistance%20Enhance%20or%20Erode%20Expertise%20Evidence%20from%20a%20Three-Month%20Field%20Experiment%20in%20Patent%20Drafting.pdf

[3] AEA RCT Registry. "AI & Expertise research with patent lawyers." AEARCTR-0015823（2025 年 5 月 21 日公開）. 【補助調査に基づく】 https://www.socialscienceregistry.org/trials/15823

[4] Autor, D., & Rodchenko, T. (2026 年 10 月 7 日). "Does better work always mean better workers?" Google Research Blog. 【本文確認】 https://research.google/blog/does-better-work-always-mean-better-workers/

[5] Dolan, E. W. (2026 年 11 月 7 日). "Prior expertise may determine whether AI builds skills or substitutes for them, field experiment suggests." PsyPost. 【本文確認】 https://www.psypost.org/prior-expertise-may-determine-whether-ai-builds-skills-or-substitutes-for-them-field-experiment-suggests/

[6] TechFeed (2026 年 11 月 8 日).「AI はベテランの判断力を鍛え、若手の思考を代替しうる — 特許弁護士 133 人・3 ヶ月のランダム化比較試験が示した分岐点」. TechFeed 公式（TechFeed Advance チャンネル）による PsyPost 記事[5]の要約紹介. 【本文確認】 https://techfeed.io/entries/6aefaa6649a7d098b464cd7b

[7] Fortune (2026 年 10 月 10 日). David Autor 氏への取材記事（NBER 論文、"illusion of competence"）. 【本文確認】 https://fortune.com/2026/10/10/david-autor-china-shock-nber-google-paper-illusion-of-competence-gen-z/

[8] Crouch, D. (2026 年 9 月). "Who Keeps the Skill? A Randomized Trial of AI Patent Drafting at Eleven Firms." Patently-O. 【会員限定部分を除く冒頭のみ確認】 https://patentlyo.com/patent/2026/09/who-keeps-the-skill-a-randomized-trial-of-ai-patent-drafting-at-eleven-firms.html

[9] Cowen, T. (2026 年 9 月). "Does AI assistance enhance or erode expertise?" Marginal Revolution. 【本文確認】 https://marginalrevolution.com/marginalrevolution/2026/09/does-ai-assistance-enhance-or-erode-expertise.html

[10] RohitAI. Google/Autor 特許弁護士研究の解説記事. 【本文確認】 https://rohitai.com/blog/google-ai-assistance-professional-expertise-patent-lawyer-study

[11] The Decoder. "AI makes junior patent attorneys instantly better but may quietly stunt their professional growth." 【見出しの参照にとどまる】 https://the-decoder.com/ai-makes-junior-patent-attorneys-instantly-better-but-may-quietly-stunt-their-professional-growth/

[12] 野崎秀吾 (2026 年 9 月). 本論文の検出力・Type M 誤差に関する note 記事. 【補助調査に基づく】 https://note.com/shugo/n/nd2b60b059e98

[13] PyTorch Korea User Group（discuss.pytorch.kr）. Google ブログと NBER WP 35720 の解説記事（韓国語）. 【検索結果の抜粋で確認】 https://discuss.pytorch.kr/t/google-nber-133-ai/12134

[14] Brynjolfsson, E., Li, D., & Raymond, L. "Generative AI at Work." Quarterly Journal of Economics (2025). NBER Working Paper No. 31161. 【書誌情報確認】 https://econpapers.repec.org/RePEc:nbr:nberwo:31161

[15] Dell'Acqua, F., McFowland, E., Mollick, E., et al. (2023). "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality." Harvard Business School Working Paper 24-013. 【数値は二次要約に基づく】 https://empirical-economics-of-ai.vercel.app/papers/DellAcquaMcFowlandMollickEtAl_2023_NavigatingJaggedTechnological

[16] Bastani, H., et al. (2025). "Generative AI without guardrails can harm learning: Evidence from high school mathematics." PNAS, 122(26), e2422633122. 【本文確認】 https://www.pnas.org/doi/10.1073/pnas.2422633122

[17] Otis, N., Clarke, R., Delecourt, S., Holtz, D., & Koning, R. "The Uneven Impact of Generative AI on Entrepreneurial Performance." Management Science (2026 年 7 月). DOI: 10.1287/mnsc.2024.06909. 【要旨確認】 https://www.researchgate.net/publication/377475913_The_Uneven_Impact_of_Generative_AI_on_Entrepreneurial_Performance

[18] Lee, H.-P., Sarkar, A., et al. (2025). "The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers." CHI '25. DOI: 10.1145/3706598.3713778. 【要旨・二次要約に基づく】 https://doi.org/10.1145/3706598.3713778

[19] Kosmyna, N., et al. (2025). "Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task." MIT Media Lab（査読前）. 【要旨・二次要約に基づく】 https://www.media.mit.edu/publications/your-brain-on-chatgpt/

[20] Choi, J. H., & Schwarcz, D. "AI Assistance in Legal Analysis: An Empirical Study." Journal of Legal Education, Vol. 73. SSRN abstract 4539836. 【要旨に基づく。URL 未記載】

[21] Choi, J. H., Monahan, A., & Schwarcz, D. (2024). "Lawyering in the Age of Artificial Intelligence." Minnesota Law Review, Vol. 109. 【要旨に基づく。URL 未記載】

[22] Bednar ほか (2026). Northwestern University Law Review 掲載予定論文（法学生約 100 人の AI 利用実験）. SSRN abstract 6525800. 【題名未確認。内容は Legal Theory Stack（L. Solum）経由の二次情報】 https://ssrn.com/abstract=6525800

[23] 日本弁理士会 (2025 年 4 月). 「弁理士業務 AI 利活用ガイドライン」. 【本文確認】
https://www.jpaa.or.jp/cms/wp-content/uploads/2025/04/AIservices-guideline.pdf

[24] 東京弁護士会 (2026). 会報「LIBRA」2026 年 7・8 月号 特集記事. 【本文確認】
https://www.toben.or.jp/message/libra/pdf/2026_0708/P02-18.pdf

[25] 松尾剛行. AI と若手法曹の育成（OJT）に関する note 記事. 【本文確認】
https://note.com/matsuo1984/n/n211b527dc2d0