ChatGPT

Curadoria de Livros e Artigos sobre IA Generativa para Consultoria

Livros estratégicos (consultoria de transformação digital e estratégia em IA)

- **Rewired: The McKinsey Guide to Outcompeting in the Age of Digital and AI** – *Eric Lamarre, Kate Smaje, Rodney Zemmel (McKinsey & Company)* – Link – Manual **prático** de transformação digital e IA usado pela McKinsey, com playbooks e casos reais que mostram como "**reprogramar**" a empresa para competir na era digital [1] [2], indispensável para consultores formularem estratégias de mudança organizacional sustentáveis.

- **Competing in the Age of AI: Strategy and Leadership When Algorithms and Networks Run the World** – *Marco Iansiti, Karim R. Lakhani (Harvard Business Review Press)* – Link – Explora como **modelos de negócios baseados em IA** rompem limitações tradicionais, redesenhando operações e estratégia empresarial [3]; leitura essencial para entender como dados e algoritmos podem redefinir vantagens competitivas e liderança corporativa.

- **All-In on AI: How Smart Companies Win Big with Artificial Intelligence** – *Thomas H. Davenport, Nitin Mittal (Harvard Business Review Press)* – Link – Destaca empresas que **abraçaram totalmente a IA**, oferecendo insights estratégicos, casos de uso e melhores práticas para gerar valor em larga escala; mostra aos consultores como liderança e cultura "all in" em IA podem impulsionar **inovação e vantagem competitiva** [4].

- **Human + Machine: Reimagining Work in the Age of AI** – *Paul R. Daugherty, H. James Wilson (Harvard Business Review Press)* – Link – Defende a **colaboração homem-máquina** na empresa, ilustrando através de exemplos como integrar IA aos processos de trabalho para multiplicar a inovação e a produtividade [5]; traz um framework prático ("Missing Middle") que ajuda consultores a repensar papéis humanos na era da IA em vez de focar apenas em automação.

- **Power and Prediction: The Disruptive Economics of Artificial Intelligence** – *Ajay Agrawal, Joshua Gans, Avi Goldfarb (Harvard Business Review Press)* – Link – Mostra por que muitas organizações **falham em capturar valor** da IA e argumenta que o verdadeiro potencial da IA surge ao redesenhar sistemas e processos inteiros – não apenas automatizar tarefas isoladas [6]; leitura valiosa para consultores entenderem como superar a "lacuna do piloto" e alinhar estratégia, dados e processos para tirar proveito das previsões em escala.

Whitepapers técnicos e empresariais (arquiteturas, padrões, riscos e estratégias de GenAI)

- **From Experiments to Deployments: A Practical Path to Scaling AI** – *OpenAI (2025)* – PDF – Guia empresarial da OpenAI que apresenta um **roteiro em quatro fases** para escalar IA nas organizações (da fundação e governança até a fluência em IA e produto em escala), com passos concretos para ir de pilotos a soluções de AI integradas no negócio [7]; é relevante para consultores pois aborda como estruturar governance, dados e times para superar o estágio de experimentação e gerar impacto real.

- **A Platform Approach to Scaling Generative AI in the Enterprise** – *Google Cloud (2023)* – Whitepaper – Apresenta a visão do Google de que escalar GenAI exige uma **plataforma integrada**, não apenas modelos isolados, detalhando arquiteturas de referência na nuvem, boas práticas de MLOps e governança para implementar IA generativa em larga escala nas empresas [8]; ajuda consultores técnicos a entender padrões para entregar soluções robustas com rapidez e segurança.

- **Microsoft Guide for Securing the AI-Powered Enterprise (Issue 1: Getting Started with AI Applications)** – *Microsoft (2025)* – Whitepaper – Primeiro de uma série de guias da Microsoft sobre **segurança e governança em IA**, cobrindo como identificar e mitigar riscos de aplicações com IA generativa (ex.: vazamento de dados, ataques de prompt injection, falhas de agentes autônomos) em conformidade com novas regulações [9]; leitura importante para consultores definirem políticas de *AI governance* e **guardrails** ao implementar soluções de GenAI de forma responsável.

- **Building Trusted AI in the Enterprise** – *Anthropic (2024)* – PDF – Guia da Anthropic que, apoiado em milhares de implantações do Claude, ensina como **escalar IA de forma confiável** na empresa, enfatizando identificação de casos de uso de alto impacto, bases sólidas (dados, segurança) e então expansão gradual do que funciona [10]; é relevante para consultores pois combina melhores práticas e lições reais para equilibrar inovação rápida com princípios de confiança, transparência e alinhamento ético.

- **The Economic Potential of Generative AI: The Next Productivity Frontier** – *McKinsey & Company (McKinsey Global Institute, 2023)* – Relatório – Estudo aprofundado quantificando como a IA generativa pode adicionar **US$2.6 a 4.4 trilhões** em valor anual à economia [11], analisando 60+ casos de uso em funções de negócio e estimando impactos em setores e no futuro do trabalho; armam consultores estratégicos com dados concretos sobre onde a GenAI traz mais impacto e quais desafios de produtividade e organização esperar.

- **AI's Trillion-Dollar Opportunity (Tech Report 2024)** – *Bain & Company (2024)* – Relatório – Destaca que a **corrida da IA** está acelerando: o mercado de produtos e serviços de IA deve alcançar até ~$900 bilhões em 2027 [12], impulsionado principalmente pelos hyperscalers (cloud e modelos fundação) e por inovações em modelos menores e software; ajuda consultores a entender tendências de investimento, novas oportunidades em infra-estrutura (como demanda por GPUs/DPUs) e como empresas devem se posicionar para capturar valor nessa expansão.

- **The CEO's Guide to the Generative AI Revolution** – *Boston Consulting Group (2023)* – Artigo – Orientação da BCG para alta liderança sobre a revolução da IA generativa, discutindo implicações

estratégicas em nível de **modelo de negócio** e organização. O texto alerta que o avanço da GenAI pode **destruir ou criar vantagens competitivas** em quase todo setor, exigindo ação deliberada dos CEOs [13] – em suma, fornece aos consultores argumentação e linguagem acessível para engajar executivos na definição de uma estratégia clara de IA.

- **Generative AI Lens – AWS Well-Architected Framework** – *Amazon Web Services (2023)* – Whitepaper – Extensão do framework de arquitetura da AWS voltada a aplicações de IA generativa, cobrindo **boas práticas** de design em nuvem (segurança, confiabilidade, custo, eficiência) para workloads com modelos fundação. Inclui diretrizes específicas como implementação de guardrails contra respostas tóxicas, padrões de RAG, gerenciamento de dados e desempenho escalável [14]; auxilia consultores e arquitetos a desenhar soluções GenAI robustas seguindo padrões validados de mercado.

- **Generative AI Architecture Patterns** – *Databricks (2023)* – Artigo técnico – Apresenta os **quatro padrões arquiteturais** principais para soluções com LLM em empresas – *Prompt Engineering*, *RAG (Retrieval-Augmented Generation)*, *Fine-Tuning* e *Pretraining* – explicando quando usar cada abordagem e como combiná-las [15]. Útil para consultores técnicos entenderem o leque de alternativas de implementação em projetos de GenAI, considerando custo, qualidade e rapidez de cada estratégia, com ênfase em ferramentas do ecossistema Databricks (Lakehouse, MosaicML).

- **Palantir Artificial Intelligence Platform (AIP) – Overview** – *Palantir (2023)* – Documento – Descreve a plataforma de IA da Palantir voltada a conectar LLMs aos dados e operações corporativas de forma segura. O AIP integra IA generativa aos processos de negócio existentes, oferecendo estúdio de agentes, recursos de observabilidade, avaliações e **governança integrada** para garantir compliance e auditabilidade [16] [17]; relevante para consultores entenderem como soluções de IA generativa podem ser implementadas ponta-a-ponta em ambientes complexos (*enterprise*), atendendo requisitos de TI e regulatórios sem deixar de entregar automação operacional via agentes de IA.

Papers acadêmicos aplicados (boas práticas, padrões e armadilhas na construção de soluções com LLMs)

- **The Prompt Report: A Systematic Survey of Prompt Engineering Techniques** – *Schulhoff et al. (2024/25)* – Paper – **Panorama abrangente** de técnicas de *prompting* para LLMs, organizando a literatura dispersa em uma taxonomia com 58 métodos e 33 termos-chave, além de diretrizes e melhores práticas para engenharia de prompts [18]. É valioso para consultores (especialmente técnicos) pois compila num só lugar tudo que se sabe sobre elaborar prompts eficazes – desde padrões simples até estratégias avançadas – ajudando a melhorar resultados de LLMs de forma sistemática em aplicações reais.

- **ReAct: Synergizing Reasoning and Acting in Language Models** – *Yao et al. (Google/Princeton, ICLR 2023)* – Paper – Introduz o paradigma **Reason+Act (ReAct)**, que ensina LLMs a gerar cadeias de raciocínio *e* ações (como consultas a ferramentas) de forma intercalada [19]. Essa abordagem mostrou-se fundamental para construir *agents* mais eficazes – por exemplo, um chatbot que pensa em passos e faz buscas externas antes de responder – melhorando interpretabilidade e controle. Consultores técnicos podem aplicar ReAct como padrão de projeto ao criar agentes de linguagem

que interajam com sistemas corporativos, evitando tanto alucinações quanto decisões sem justificativa.

- **A Survey on RAG Meeting LLMs: Towards Retrieval-Augmented Large Language Models** – *Fan et al. (KDD 2024)* – Paper – Revisão completa sobre métodos de **Geração Aumentada por Recuperação (RAG)** em LLMs, cobrindo arquiteturas, técnicas de treinamento e aplicações práticas. Discute como combinar modelos de linguagem com bases de conhecimento externas para mitigar limitações como alucinações e conhecimento desatualizado [20]. Para consultores, o paper oferece um **guia atualizado** de como implementar RAG (ex.: chatbots com busca documental) e alerta sobre desafios técnicos a considerar – fundamental dado o interesse crescente de clientes em *LLMs conectados aos seus dados*.

- **Holistic Evaluation of Language Models (HELM)** – *Liang et al. (Stanford CRFM, TMLR 2023)* – Paper – Propõe um framework aberto para **avaliar LLMs de forma abrangente**, cobrindo uma ampla gama de cenários de uso e múltiplas métricas de performance e risco (exatidão, robustez, calibração, vieses, toxicidade, eficiência) [21]. O HELM já benchmarkeou dezenas de modelos sob condições padronizadas, expondo pontos fortes e fraquezas de cada um. Consultores (perfil técnico ou estratégico) podem usar insights do HELM para entender limitações dos modelos em diferentes tarefas, orientar a escolha de modelos para cada caso e conscientizar clientes sobre questões de **viés e qualidade** que precisam monitorar nas soluções de IA.

- **Constitutional AI: Harmlessness from AI Feedback** – *Bai et al. (Anthropic, 2022)* – Paper – Apresenta um método inovador de **alinhamento de LLMs** onde a própria IA se auto-refina seguindo um conjunto de princípios ("Constituição") pré-definidos, em vez de depender exclusivamente de feedback humano. O processo envolve o modelo gerar autocríticas e ajustes de suas respostas e depois aplicar *reinforcement learning* com uma recompensa dada por outro modelo crítico, resultando numa IA que **recusa pedidos inadequados de forma não-evasiva** e explica suas recusas [22] [23]. Este paper é relevante para consultores pois indica caminhos para implementar *guardrails* éticos e de segurança em assistentes de IA corporativos reduzindo esforço de supervisão humana – um grande desafio em soluções GenAI empresariais (ex.: evitar respostas tóxicas ou inseguras em escala).

---

[1] [2] Rewired: The McKinsey Guide to Outcompeting in the Age of Digital and AI by Eric Lamarre, Kate Smaje, Rodney Zemmel, Hardcover | Barnes & Noble®
https://www.barnesandnoble.com/w/rewired-eric-lamarre/1143372229

[3] Best Essential AI Alternatives | AI Tech Suite
https://www.aitechsuite.com/alternatives/essential.ai

[4] All-in On AI: Tom Davenport's New Book - First Analytics
https://firstanalytics.com/all-in-on-ai-tom-davenports-new-book/

[5] Human + Machine, Updated and Expanded: Reimagining Work in ...
https://store.hbr.org/product/human-machine-updated-and-expanded-reimagining-work-in-the-age-of-ai/10724?srsltid=AfmBOorVmELpksjD3YaHn49R3lvijmaFJpBuzfySFWS4oIip6FcwDlOi

[6] Power and Prediction by Ajay Agrawal, Joshua Gans, Avi Goldfarb ...
https://www.summrize.com/books/power-and-prediction-summary

7 Practical Path + People = AI Outcomes
https://www.linkedin.com/pulse/practical-path-people-ai-outcomes-tim-creasey-ce6gc

8 Google Cloud Whitepapers | Google Cloud
https://cloud.google.com/whitepapers

9 Securing AI: Navigating risks and compliance for the future | The Microsoft Cloud Blog
https://www.microsoft.com/en-us/microsoft-cloud/blog/2025/04/23/securing-ai-navigating-risks-and-compliance-for-the-future/

10 assets.anthropic.com
https://assets.anthropic.com/m/66daaa23018ab0fd/original/Anthropic-enterprise-ebook-digital.pdf

11 Economic potential of generative AI | McKinsey
https://www.mckinsey.com/capabilities/tech-and-ai/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier

12 AI's Trillion-Dollar Opportunity | Bain & Company
https://www.bain.com/insights/ais-trillion-dollar-opportunity-tech-report-2024/

13 The CEO's Guide to the Generative AI Revolution | BCG
https://www.bcg.com/publications/2023/ceo-guide-to-ai-revolution

14 AWS Generative AI Best Practices Framework v2
https://docs.aws.amazon.com/audit-manager/latest/userguide/aws-generative-ai-best-practices.html

15 Generative AI Architecture Patterns | Databricks
https://www.databricks.com/product/machine-learning/build-generative-ai

16 17 Overview • AIP • Palantir
https://www.palantir.com/docs/foundry/aip/overview

18 [2406.06608] The Prompt Report: A Systematic Survey of Prompt Engineering Techniques
https://arxiv.org/abs/2406.06608

19 ReAct: Synergizing Reasoning and Acting in Language Models
https://research.google/blog/react-synergizing-reasoning-and-acting-in-language-models/

20 [2405.06211] A Survey on RAG Meeting LLMs: Towards Retrieval-Augmented Large Language Models
https://arxiv.org/abs/2405.06211

21 [2211.09110] Holistic Evaluation of Language Models
https://arxiv.org/abs/2211.09110

22 23 [2212.08073] Constitutional AI: Harmlessness from AI Feedback
https://arxiv.org/abs/2212.08073