Quais são hoje os principais usos para GenAI, considerando não apenas aquilo que aprendemos nos primeiros anos dessa tecnologia, mas principalmente aquilo em que ela se transformou em um cenário no qual modelos generativos passaram a fazer parte de aplicações corporativas, plataformas de produtividade, ambientes de desenvolvimento, ferramentas de atendimento, soluções analíticas e, cada vez mais, sistemas capazes de executar ações de maneira parcialmente autônoma sobre processos empresariais?

Creio que essa é uma questão cada vez mais comum nas empresas, seja para evoluir seu próprio patamar de inovação e eficiência operacional, seja para compreender aquilo que concorrentes já estão fazendo e reduzir o risco de ficar para trás em uma

transformação que claramente deixou de ser apenas experimental. O debate já não gira apenas em torno de chatbots capazes de escrever textos ou responder perguntas, pois GenAI começou a funcionar como uma camada tecnológica presente em diferentes partes da organização, participando diretamente da maneira como pessoas trabalham, software é desenvolvido, conhecimento é acessado, decisões são apoiadas e determinados processos são executados.

Aqui deixo a recomendação de um webinar com uma visão bastante estruturada do Gartner sobre aplicações e casos de uso de GenAI:

https://webinar.gartner.com/524786/agenda/session/1184390

O mundo da tecnologia está em constante evolução, e a Inteligência Artificial assumiu um papel cada vez mais significativo nessa transformação, mas talvez poucas tecnologias recentes tenham mudado tão rapidamente de natureza quanto a GenAI. Em um intervalo relativamente curto, passamos de uma fase marcada principalmente pela geração de textos e imagens para um cenário muito mais amplo, no qual modelos conseguem trabalhar de maneira multimodal, analisar documentos extensos, interpretar imagens e vídeos, compreender voz, escrever código, utilizar ferramentas externas e participar de workflows empresariais completos.

Essa evolução muda bastante a discussão sobre casos de uso, porque o foco inicial estava muito concentrado em produtividade individual, geração de conteúdo, resumo de documentos e assistentes conversacionais, enquanto hoje a oportunidade começa a avançar para workflows de negócio nos quais AI Agents conseguem consultar informações, interpretar contexto, acessar sistemas autorizados, executar múltiplas etapas e devolver resultados sem que uma pessoa precise conduzir manualmente cada ponto do processo.

E vale reiterar mais uma vez que considero tudo isso apenas o começo, não no sentido de repetir o entusiasmo genérico que costuma acompanhar qualquer nova tecnologia, mas porque ainda estamos observando uma velocidade extraordinária de evolução simultaneamente em modelos, infraestrutura, multimodalidade, custo, context windows, reasoning, tool use, integração com sistemas e agentes capazes de operar com níveis crescentes de autonomia.

Talvez a principal mudança de perspectiva esteja justamente aí, pois a pergunta já não deveria ser apenas “o que a GenAI consegue gerar?”, mas “qual trabalho ela consegue ajudar a realizar, quais decisões consegue apoiar, quais atividades pode executar e qual resultado empresarial pode efetivamente produzir?”. Essa diferença parece simples, mas muda completamente a maneira como uma organização deveria estruturar sua estratégia de AI.

Ao mesmo tempo, quanto maior a capacidade da tecnologia de agir e não apenas responder, mais importantes se tornam questões relacionadas a segurança, privacidade, compliance, identidade, qualidade dos dados, governança e accountability. Um modelo que gera um texto incorreto possui determinado nível de risco, enquanto um agente com acesso a aplicações, dados ou transações pode transformar um erro de raciocínio em uma ação concreta, ampliando significativamente as consequências potenciais.

Creio que a fase atual da GenAI dentro das empresas pode ser descrita como uma transição da experimentação para a industrialização, na qual a tecnologia já demonstrou capacidade suficiente para que a discussão deixe de ser puramente hipotética, mas capturar valor em escala exige muito mais do que disponibilizar uma interface de chat para milhares de usuários. Exige escolher corretamente os casos de uso, redesenhar processos, preparar dados, integrar plataformas, estabelecer governança, medir resultados e, principalmente, compreender onde a tecnologia efetivamente cria valor em vez de simplesmente adicionar uma nova camada de complexidade.

O artigo do Gartner

O webinar do Gartner utilizado como ponto de partida para esta reflexão aborda a expansão das capacidades dos sistemas de Inteligência Artificial e, particularmente, o potencial da IA generativa para criar conteúdo, dados sintéticos, software, mídia e novas formas de conhecimento a partir de grandes volumes de informação. Algumas das previsões apresentadas naquele contexto já pertencem ao passado e, por isso, considero hoje mais útil observar aquilo que efetivamente aconteceu desde então do que continuar tratando essas estimativas como expectativas futuras.

A geração assistida de código, por exemplo, deixou de ser uma curiosidade tecnológica e passou a fazer parte do workflow de uma parcela significativa dos desenvolvedores, enquanto dados sintéticos, geração multimodal, copilots e interfaces conversacionais foram incorporados a diferentes produtos empresariais. Ao mesmo tempo, aquilo que inicialmente parecia apenas um novo modelo de interação passou a evoluir para arquiteturas em que modelos conseguem participar de fluxos de decisão e execução.

O ponto central do estudo, entretanto, continua extremamente atual, porque GenAI não representa apenas mais uma categoria de software, mas uma tecnologia horizontal capaz de afetar praticamente qualquer processo que envolva linguagem, conhecimento, análise, criação, decisão ou interação digital. Isso explica por que os casos de uso se multiplicaram de maneira tão rápida e começaram a aparecer em

praticamente todas as funções empresariais.

Em Marketing, GenAI pode produzir variações de conteúdo, campanhas, imagens, vídeos e personalização em escala, enquanto em Software Engineering pode apoiar escrita, revisão, documentação e testes de código. Em Customer Service, pode resumir históricos, sugerir respostas, consultar bases de conhecimento e resolver determinadas solicitações de ponta a ponta, enquanto em áreas corporativas pode apoiar análise de contratos, pesquisa, criação de documentos, síntese de reuniões, preparação de apresentações e processamento de informações.

Em Data & Analytics, a tecnologia começa a permitir que usuários interajam com informações através de linguagem natural, enquanto em Cybersecurity pode apoiar investigação, triagem, análise e resposta a incidentes. Essa lista continua crescendo justamente porque GenAI está deixando de ser uma aplicação específica e se transformando em uma capability incorporada a diferentes plataformas.

De Generative AI para Agentic AI

Uma das mudanças mais relevantes dos últimos anos está na transição entre GenAI predominantemente conversacional e Agentic AI, na qual sistemas deixam de apenas responder a perguntas e passam a executar sequências de ações com determinado nível de autonomia. Um chatbot tradicional pode explicar como determinada atividade deveria ser realizada, enquanto um agente pode potencialmente localizar as informações necessárias, acessar ferramentas autorizadas, executar determinadas etapas, verificar o resultado e concluir a tarefa.

Essa mudança desloca o valor da tecnologia do simples acesso ao conhecimento para a execução de trabalho, aproximando AI de processos empresariais que anteriormente dependiam integralmente de pessoas para conectar cada etapa. O ponto mais importante não está em imaginar uma substituição generalizada do trabalho humano, mas em compreender que tarefas antes distribuídas entre diferentes sistemas e interfaces podem começar a ser orquestradas por uma camada inteligente.

Do conhecimento para a execução

Quando um modelo apenas apresenta conhecimento, o usuário continua responsável por interpretar a resposta e transformar aquela informação em ação, enquanto em um workflow agentic parte desse trabalho pode ser executada diretamente pelo sistema, reduzindo a distância entre intenção e resultado. Essa mudança pode parecer apenas operacional, mas altera profundamente o potencial de automação de processos intensivos em conhecimento.

Em vez de funcionar apenas como uma espécie de consultor digital que fornece orientação, a AI pode começar a assumir responsabilidades específicas dentro de uma cadeia de execução, especialmente em atividades nas quais as regras, ferramentas disponíveis e limites de atuação estejam suficientemente definidos.

Da resposta única para processos de múltiplas etapas

Um agente pode decompor uma atividade em subtarefas, selecionar ferramentas diferentes para cada etapa, consultar novas informações conforme necessário e ajustar sua trajetória ao longo da execução, aproximando-se muito mais da maneira como determinados processos humanos realmente funcionam. Isso permite pensar em AI não apenas como uma interface, mas como parte da própria arquitetura de execução.

Essa capacidade se torna particularmente interessante em processos nos quais uma pessoa precisa navegar entre diversos sistemas, consultar informações em múltiplas fontes e consolidar o resultado final, pois exatamente essa combinação de interpretação, pesquisa e ação começa a ser absorvida por arquiteturas agentic.

Da produtividade individual para produtividade operacional

Copilots aumentam principalmente a capacidade de uma pessoa, enquanto agents possuem potencial para modificar a própria estrutura operacional de um processo, eliminando determinadas atividades manuais e permitindo que profissionais atuem prioritariamente sobre exceções, decisões e situações complexas. Essa diferença é particularmente importante porque o ganho deixa de depender exclusivamente de cada usuário conseguir utilizar bem uma ferramenta e passa a ser incorporado ao próprio workflow.

A consequência é que AI começa a impactar não apenas produtividade individual, mas também desenho organizacional, capacidade operacional e distribuição de trabalho, criando discussões que envolvem Technology, Operations, HR, Risk e diferentes áreas de negócio.

Essa evolução também explica por que o tema de governance se tornou muito mais importante, pois um modelo que escreve um e-mail inadequado pode produzir constrangimento, mas um agente autorizado a modificar dados, aprovar transações ou executar operações exige controles muito mais sofisticados sobre identidade, autorização, limites de ação, logs, observability e mecanismos de interrupção.

Os principais casos de uso mudaram de

natureza

No início da adoção corporativa, era relativamente comum organizar os casos de uso de GenAI em categorias simples como geração de texto, resumo, pesquisa, tradução, criação de imagens e desenvolvimento de software, e todas essas categorias continuam válidas. O que mudou é que elas passaram a representar apenas os primeiros níveis de utilização, enquanto aplicações mais sofisticadas começam a combinar várias dessas capacidades dentro do mesmo processo.

Acredito que hoje exista uma maneira mais útil de estruturar a discussão, olhando para o tipo de valor que a tecnologia entrega e para a profundidade de sua participação no trabalho. Em vez de pensar apenas no formato do output, podemos analisar o quanto a AI participa da atividade e qual responsabilidade permanece com o ser humano.

Assistência

Nesse primeiro nível, AI ajuda uma pessoa a executar uma atividade que continuará essencialmente humana, seja escrevendo um documento, resumindo informações, preparando uma apresentação, analisando um contrato ou criando uma primeira versão de determinado conteúdo. O ganho principal aparece na redução do esforço necessário para começar, estruturar ou revisar determinada tarefa.

Esse tipo de utilização tende a apresentar adoção relativamente simples porque não exige necessariamente redesenho profundo do processo, permitindo que o usuário continue responsável pela decisão final e utilize a tecnologia como mecanismo de aceleração.

Augmentation

Em um nível mais sofisticado, a tecnologia passa a ampliar capacidade humana através de análise, recomendações, recuperação de conhecimento, simulações ou geração de alternativas, permitindo que o profissional execute atividades mais complexas ou tome decisões com maior quantidade de informação. Nesse caso, a contribuição da AI está menos relacionada à simples velocidade e mais à qualidade e amplitude do raciocínio disponível.

Isso pode ocorrer, por exemplo, quando um profissional recebe diferentes cenários para determinada decisão, quando um analista consegue consultar uma grande quantidade de documentos simultaneamente ou quando um especialista utiliza AI para explorar hipóteses que exigiriam muito mais tempo manualmente.

Automation

Determinadas atividades podem ser executadas quase integralmente pela tecnologia quando são suficientemente estruturadas, repetitivas e possuem mecanismos claros de validação. A importância dessa categoria está no fato de que produtividade deixa de depender apenas de cada pessoa trabalhar mais rápido e passa a ser incorporada ao próprio processo.

Quanto maior o nível de automação, maior também precisa ser a preocupação com exception handling, qualidade dos dados, limites operacionais e mecanismos de fallback, pois o objetivo não deveria ser apenas eliminar intervenção humana, mas garantir que o processo continue funcionando de maneira segura quando situações imprevistas ocorrerem.

Autonomy

Finalmente, Agentic AI cria a possibilidade de sistemas que recebem um objetivo e determinam parcialmente como alcançá-lo, utilizando ferramentas, dados e processos dentro de limites estabelecidos. Esse nível pode produzir ganhos extraordinários, mas também exige muito mais cuidado porque a organização está delegando não apenas execução, mas uma parcela do próprio raciocínio operacional.

A autonomia não deveria ser tratada como objetivo automático para todos os processos, mas como uma escolha de arquitetura e governance que precisa ser proporcional ao risco, à reversibilidade das ações e à capacidade de monitoramento disponível.

Esses quatro níveis não representam necessariamente uma sequência obrigatória, porque determinados processos permanecerão permanentemente no modelo de assistência enquanto outros poderão avançar rapidamente para automação ou autonomia. A decisão correta depende muito mais do contexto, do risco e do valor esperado do que simplesmente da capacidade técnica disponível.

Reflexões sobre a implementação de GenAI

A adoção de GenAI representa uma transformação importante no panorama tecnológico atual, mas aquilo que inicialmente parecia uma nova categoria de aplicação SaaS evoluiu para um ecossistema muito mais complexo, envolvendo modelos fundacionais, APIs, plataformas de AI, serviços de Cloud, modelos especializados, componentes open source, infraestrutura privada e diferentes mecanismos de integração.

Por isso, algumas reflexões estratégicas continuam sendo especialmente importantes

para empresas que desejam capturar valor sem transformar entusiasmo tecnológico em um portfolio descontrolado de experimentos, principalmente porque a facilidade para iniciar um piloto continua muito maior do que a capacidade necessária para industrializá-lo com segurança.

Generative AI será majoritariamente consumida como serviço, mas não exclusivamente

A maior parte das empresas provavelmente continuará utilizando modelos e plataformas fornecidos por grandes players, porque treinar e operar modelos fundacionais competitivos exige volumes extraordinários de capital, infraestrutura, energia, dados e talentos especializados. Isso não significa, entretanto, que toda AI empresarial dependerá exclusivamente de serviços públicos ou de um único provider.

Hoje existem arquiteturas híbridas nas quais diferentes modelos podem ser selecionados conforme requisitos de custo, latência, segurança, soberania, qualidade e tipo de workload, incluindo modelos comerciais, open-weight, especializados e soluções executadas sobre infraestrutura privada. Dessa forma, a discussão estratégica passa muito mais por model portfolio management do que por escolher simplesmente um único fornecedor para tudo.

O impacto econômico para os vendors é gigantesco

Os grandes fornecedores de Cloud e software corporativo estão incorporando AI praticamente em todo o portfolio, enquanto novos players construíram negócios inteiros ao redor de modelos fundacionais, agentes, infraestrutura, observability e ferramentas de desenvolvimento. Isso cria um mercado gigantesco, mas também aumenta a possibilidade de dependência tecnológica.

Se processos empresariais críticos começarem a utilizar agentes, memória, tools, proprietary APIs e serviços específicos de determinada plataforma, o custo de mudança poderá crescer rapidamente, tornando arquitetura e procurement igualmente relevantes na estratégia de AI. A discussão sobre vendor lock-in, portanto, ganha uma nova dimensão quando a lógica operacional passa a ser construída ao redor do comportamento de modelos e agentes.

O risco competitivo é real, mas a ideia de “evento de extinção” precisa ser utilizada com cuidado

Continuo acreditando que ignorar AI pode produzir desvantagem competitiva relevante em diversos setores, principalmente quando concorrentes conseguirem operar com menor custo, maior velocidade, melhor personalização ou ciclos de inovação significativamente mais curtos. Ao mesmo tempo, considero igualmente perigoso utilizar esse argumento para justificar qualquer investimento.

O fato de AI ser estratégica não significa que qualquer projeto de AI seja estratégico, e a verdadeira diferenciação provavelmente surgirá menos da simples utilização da tecnologia e muito mais da capacidade de aplicá-la a processos, dados e conhecimento específicos da empresa. O acesso ao modelo pode ser amplamente democratizado, mas a forma como ele é integrado ao negócio continua sendo particular.

Custos precisam ser analisados desde o início

Uma das grandes armadilhas da primeira onda de GenAI foi a percepção de que uma barreira de entrada relativamente baixa significaria automaticamente economics simples, quando na prática existe enorme diferença entre criar um protótipo utilizado por poucas pessoas e operar centenas de agentes executando milhares de atividades todos os dias.

Custos podem envolver consumo de modelos, infraestrutura, armazenamento, ferramentas, integrações, observability, segurança, implementação, suporte, evolução e equipes especializadas, fazendo com que Business Case e FinOps for AI ganhem importância crescente. A organização precisa compreender a economics do workload antes que o sucesso técnico do piloto se transforme em uma surpresa financeira quando a escala chegar.

As lições de Cloud continuam extremamente úteis

Durante a primeira grande onda de Cloud, diversas empresas descobriram que simplesmente mover workloads sem redesenhar arquitetura raramente capturava todo o valor possível, e com AI existe um risco semelhante. Adicionar um Copilot a um processo ruim pode tornar determinadas etapas mais rápidas, mas dificilmente produzirá a mesma transformação que redesenhar o processo considerando desde o início aquilo que AI permite automatizar, simplificar ou eliminar.

Essa talvez seja uma das maiores lições para os próximos anos, porque não basta “adicionar AI” a cada sistema ou atividade existente. É preciso repensar o trabalho, questionar etapas que deixaram de ser necessárias e utilizar a tecnologia para simplificar a operação em vez de apenas acelerar aquilo que já era feito.

GenAI não é apenas mais um hype tecnológico

Naturalmente, qualquer tecnologia que recebe tamanha quantidade de atenção desperta perguntas sobre hype, expectativas exageradas e sustentabilidade, principalmente porque a história da tecnologia está repleta de conceitos que pareciam destinados a transformar tudo ao seu redor e posteriormente encontraram aplicações

muito mais restritas do que inicialmente imaginado.

No caso de GenAI, entretanto, considero que algumas evidências já tornam bastante difícil tratá-la simplesmente como uma moda passageira, porque a tecnologia foi incorporada às principais plataformas de produtividade, desenvolvimento de software, Cloud, CRM, ERP, Cybersecurity, Data e Collaboration, ao mesmo tempo em que criou novas categorias de produtos e alterou a própria interface através da qual usuários interagem com sistemas digitais.

Mais importante ainda, AI está deixando de ser apenas um produto que o usuário explicitamente abre para utilizar e passa a funcionar como uma capability embutida em outros produtos, o que significa que muitas pessoas utilizarão AI sem necessariamente pensar conscientemente nisso. Esse talvez seja um sinal mais importante de maturidade do que qualquer comparação sobre quantidade de usuários de determinada ferramenta.

Quando uma tecnologia deixa de ser percebida como algo separado e passa a funcionar como parte natural da experiência, ela começa a se aproximar de um estágio muito mais profundo de adoção. Foi assim com Cloud, mobile, Search e diversas outras capabilities que inicialmente eram tratadas como categorias próprias e posteriormente passaram a fazer parte da infraestrutura invisível da vida digital.

O que é Inteligência Artificial Generativa?

A Inteligência Artificial Generativa refere-se a uma categoria de sistemas de AI capaz de produzir novos outputs a partir de padrões aprendidos em grandes volumes de informação, incluindo texto, imagens, áudio, vídeo, software, estruturas de dados e diferentes formas de conteúdo. Sua evolução recente tornou essa definição mais ampla porque os modelos passaram a trabalhar simultaneamente com múltiplas modalidades e a interagir com ferramentas externas.

Um modelo moderno pode receber texto, imagem ou áudio como entrada, analisar documentos extensos, interpretar determinado contexto, gerar diferentes tipos de conteúdo e utilizar software ou APIs para executar tarefas. Essa capacidade faz com que a fronteira entre Generative AI e outras modalidades de Artificial Intelligence fique cada vez menos rígida.

Modelos generativos podem funcionar junto a mecanismos de busca, Machine Learning tradicional, sistemas preditivos, bancos vetoriais, conhecimento corporativo, regras de negócio e outros componentes especializados, formando arquiteturas nas quais cada elemento cumpre uma função específica. A solução empresarial raramente será apenas “um LLM”, mas uma combinação de diferentes capacidades técnicas

integradas ao processo que a organização pretende transformar.

Grandes players do mercado

O mercado de Artificial Intelligence deixou de ser uma disputa relativamente simples entre alguns poucos modelos e se transformou em um ecossistema no qual diferentes fornecedores competem simultaneamente em modelos fundacionais, plataformas empresariais, developer tools, AI Agents, infraestrutura e integração com aplicações.

Considero mais útil olhar para esses players pela posição estratégica que ocupam do que simplesmente tentar identificar qual possui “o melhor modelo”, pois essa comparação muda rapidamente e diferentes workloads podem favorecer soluções completamente distintas.

OpenAI

A OpenAI ocupa posição relevante no mercado de GenAI através do ChatGPT e de sua plataforma de modelos e APIs, tendo ampliado sua atuação para ambientes de trabalho, desenvolvimento de software e execução de tarefas agentic. Sua vantagem está fortemente associada à popularização da interface conversacional e à velocidade de evolução dos produtos, enquanto os desafios típicos do ambiente enterprise continuam envolvendo governance, integração, administração, custos, segurança e escolha adequada do modelo para cada workload.

Microsoft

A principal força da Microsoft está na capacidade de distribuir AI através de um ecossistema corporativo já amplamente utilizado, incluindo produtividade, desenvolvimento, Cloud, dados e aplicações empresariais. O conceito de Copilot evoluiu progressivamente de assistentes associados a produtos individuais para uma plataforma mais ampla de interação com AI e construção de agentes, fazendo com que integração com identidade, dados e ferramentas corporativas seja uma parte central de sua proposta.

Google

O Google combina pesquisa histórica em Artificial Intelligence, infraestrutura própria, enorme presença em Cloud, Search, Workspace e Android, além de uma família de modelos Gemini que passou a funcionar como base para diferentes produtos e experiências. A empresa também avançou de maneira importante em agent platforms, multimodalidade e integração entre AI e serviços de Cloud, reforçando uma estratégia na qual modelos e infraestrutura fazem parte do mesmo ecossistema.

Anthropic

A Anthropic consolidou uma posição relevante principalmente no ambiente empresarial e entre developers, com forte presença em coding, análise de documentos e workloads nos quais qualidade de reasoning, context handling e segurança possuem importância elevada. A disponibilidade de seus modelos através de diferentes Cloud providers também amplia a possibilidade de utilização dentro de arquiteturas multicloud.

Meta

A Meta possui uma posição diferente ao investir fortemente em modelos open-weight, criando uma alternativa particularmente relevante para organizações que desejam maior controle sobre infraestrutura, customização ou deployment. Além dos modelos, a empresa continua utilizando AI intensamente em suas plataformas sociais, recomendação de conteúdo, publicidade e produtos voltados a dispositivos e experiências digitais.

Amazon Web Services

A AWS ocupa posição estratégica não apenas pela infraestrutura necessária para executar workloads de AI, mas também através de plataformas que permitem acessar diferentes modelos e construir aplicações sem necessariamente comprometer toda a arquitetura com um único fornecedor. Isso reforça uma tendência importante de model choice como parte do próprio design da solução.

IBM

A IBM continua concentrando grande parte de sua estratégia em AI empresarial, governança, integração e ambientes híbridos, procurando explorar justamente organizações nas quais requisitos regulatórios, controle, segurança e integração com sistemas existentes possuem relevância elevada.

xAI

A xAI rapidamente deixou de poder ser tratada apenas como um novo entrante experimental e passou a competir de maneira mais direta no mercado de modelos e produtos de AI, beneficiando-se da integração com o ecossistema de empresas de Elon Musk e de investimentos expressivos em infraestrutura computacional.

É importante destacar que esse mercado continua extremamente dinâmico, com novos modelos aparecendo constantemente, providers incorporando modelos concorrentes dentro das próprias plataformas e empresas começando a compreender que talvez não exista qualquer razão para selecionar uma única tecnologia de AI para toda a organização.

A estratégia multimodelo começa a fazer mais sentido

Uma das mudanças mais interessantes dessa fase de maturidade está na percepção de que diferentes modelos apresentam características distintas de qualidade, custo, latência, segurança, multimodalidade e especialização, o que abre espaço para arquiteturas nas quais cada workload utiliza a alternativa mais apropriada em vez de obrigar toda a organização a depender de um único modelo.

Uma atividade extremamente simples pode ser executada por um modelo mais barato e rápido, enquanto um problema complexo pode justificar reasoning mais sofisticado e maior custo por interação. Um workload sensível pode exigir deployment específico ou controles adicionais, enquanto uma atividade intensiva em vídeo, áudio ou outra modalidade pode favorecer um provider completamente diferente.

Esse tipo de model routing tende a se tornar progressivamente mais natural conforme organizações amadurecem sua arquitetura de AI, da mesma maneira que diferentes serviços de Cloud passaram a ser escolhidos conforme características específicas de cada workload. A consequência é que CIOs e arquitetos precisam evitar dependência desnecessária de detalhes proprietários que tornem a substituição de modelos excessivamente complexa no futuro.

Isso não significa buscar uma abstração absoluta, porque esse tipo de estratégia pode acabar eliminando justamente os diferenciais mais interessantes de determinada plataforma, mas construir conscientemente um equilíbrio entre aproveitamento de capabilities proprietárias e preservação de optionality. Em um mercado que muda tão rapidamente, flexibilidade arquitetural possui valor estratégico.

Principais tendências de mercado

A adoção de GenAI continua avançando em praticamente todos os setores, mas as tendências atuais são significativamente mais amplas do que aquelas observadas na primeira fase da tecnologia. O ponto comum entre elas está na transição de AI como ferramenta isolada para AI como parte estruturante de produtos, plataformas e workflows.

AI Agents e workflows agentic

A evolução de copilots para agentes representa provavelmente uma das tendências mais importantes, permitindo que sistemas deixem de apenas auxiliar usuários e passem a executar partes relevantes de processos empresariais. Essa mudança amplia

potencial de produtividade, mas também aumenta a necessidade de governance e observability.

Multimodalidade

Texto, imagem, áudio e vídeo passam progressivamente a fazer parte da mesma experiência, permitindo que sistemas compreendam diferentes tipos de informação e escolham a modalidade mais adequada para cada interação. Isso expande significativamente o universo de casos de uso e aproxima AI de tarefas que anteriormente exigiam soluções especializadas.

Software Engineering assistido por AI

Coding continua sendo um dos casos de uso mais maduros, mas o escopo está se expandindo da simples geração de trechos de código para planejamento, refactoring, testes, documentação, análise de repositórios e execução de tarefas completas. A discussão começa, portanto, a migrar de “AI que escreve código” para “AI que participa do software lifecycle”.

Enterprise Search e Knowledge Management

A capacidade de interagir com conhecimento corporativo através de linguagem natural cria uma nova interface para documentação, políticas, procedimentos e informação distribuída entre múltiplos sistemas. Esse tipo de aplicação pode parecer simples, mas possui enorme potencial de reduzir tempo perdido procurando informação.

Hyperpersonalization

AI permite produzir experiências muito mais contextualizadas para clientes e colaboradores, combinando informações individuais, comportamento e conteúdo gerado dinamicamente. O desafio está em equilibrar relevância, privacidade e governança sobre como esses dados são utilizados.

Synthetic Data e conteúdo sintético

Dados artificiais podem ajudar em treinamento, simulação, testes e situações nas quais dados reais são escassos, caros ou sensíveis, embora sua utilização também exija controles sobre qualidade e representatividade. Conteúdo sintético também amplia possibilidades em marketing, treinamento e comunicação.

Small Language Models e modelos especializados

Nem todo problema exige o maior modelo disponível, e modelos menores ou especializados podem apresentar vantagens relevantes de custo, velocidade, controle e privacidade. A maturidade do mercado tende a aumentar justamente conforme empresas deixam de buscar apenas o modelo mais poderoso e passam a buscar o

modelo mais adequado.

AI Governance by Design

Governance começa a deixar de ser apenas um conjunto de documentos e comitês para se transformar em controles incorporados à própria arquitetura, algo particularmente importante quando agentes possuem capacidade de agir autonomamente. Essa transição de governance documental para governance executável será fundamental para a industrialização da tecnologia.

A economia da AI entra definitivamente na pauta

Nos primeiros ciclos de experimentação, o custo de GenAI muitas vezes não representava a principal preocupação porque os volumes eram relativamente pequenos e a maior parte das iniciativas estava restrita a Proofs of Concept, assistentes internos ou grupos reduzidos de usuários. Quando aplicações começam a operar em escala, entretanto, a situação muda completamente porque pequenos custos unitários passam a ser multiplicados por grandes volumes de interações, workflows e agentes.

Um agente pode realizar diversas chamadas para concluir uma única atividade, enquanto modelos mais sofisticados podem custar significativamente mais e contextos extensos aumentam diretamente o consumo de processamento. Workloads multimodais adicionam novas necessidades de computação, memória e armazenamento, fazendo com que a economics de uma solução se torne parte inseparável de sua própria arquitetura.

Essa realidade cria a necessidade de uma disciplina econômica própria, pois não basta saber quanto determinada plataforma de AI custa no total. É necessário compreender qual parcela desse custo está associada a cada workload, qual outcome é produzido, qual alternativa poderia executar a mesma atividade e se o valor gerado supera consistentemente o custo necessário para produzi-lo.

Unit Economics

Métricas como custo por usuário, por processo, por tarefa, por atendimento resolvido ou por execução de agente tendem a se tornar tão importantes quanto métricas técnicas, porque conectam consumo diretamente ao valor gerado. Sem esse tipo de visão, uma solução pode parecer barata no nível agregado enquanto determinadas jornadas apresentam economics claramente insustentáveis.

Essa abordagem também ajuda a comparar alternativas de arquitetura e compreender se determinado caso de uso está ganhando eficiência à medida que escala ou apenas multiplicando custos de maneira proporcional ao volume.

Model Routing

Utilizar modelos diferentes conforme a complexidade de cada tarefa pode reduzir significativamente o custo sem comprometer a experiência, principalmente quando atividades simples são encaminhadas para modelos menores e workloads complexos utilizam capacidades mais sofisticadas apenas quando necessário.

Essa lógica tende a se tornar parte fundamental da arquitetura financeira de AI, especialmente em ambientes nos quais milhares ou milhões de interações precisam ser processadas continuamente.

Caching e reutilização

Evitar processamento repetido de informações idênticas também começa a fazer parte do design econômico, principalmente em workloads nos quais o mesmo contexto ou resultado pode ser reutilizado diversas vezes. Eficiência passa, portanto, a depender não apenas do modelo escolhido, mas de como toda a arquitetura foi desenhada.

Esse tipo de otimização pode parecer extremamente técnico, mas começa a possuir impacto financeiro suficiente para fazer parte do próprio Business Case.

Background Processing

Workloads que não exigem baixa latência podem eventualmente utilizar mecanismos de processamento mais econômicos, enquanto atividades críticas podem justificar maior custo para garantir resposta imediata. Isso reforça a ideia de que a arquitetura de AI passa a possuir uma dimensão financeira tão importante quanto sua dimensão técnica.

A decisão correta deixa de ser simplesmente buscar o máximo desempenho e passa a envolver a escolha do nível de performance necessário para cada tipo de atividade.

Expectativas para o futuro da GenAI

As expectativas continuam elevadas, mas acredito que a discussão amadureceu significativamente e deixou de se concentrar apenas na capacidade impressionante de produzir conteúdo semelhante ao humano. Hoje, a questão mais estratégica está na capacidade de transformar processos, redistribuir trabalho e modificar a maneira como sistemas e pessoas colaboram.

Ampliação das capacidades humanas

GenAI continuará funcionando como instrumento de augmentation, permitindo que profissionais acessem conhecimento, explorem alternativas, criem primeiras versões e analisem informações com velocidade muito superior. Esse ganho pode ser particularmente relevante quando profissionais possuem domínio suficiente do contexto para validar e aperfeiçoar aquilo que a tecnologia produz.

Nesse cenário, AI não elimina a importância da expertise, mas pode amplificá-la ao permitir que especialistas processem quantidade muito maior de informação e explorem alternativas que anteriormente seriam inviáveis pelo custo de tempo necessário.

Colaboração Human + AI

O futuro provavelmente não será simplesmente de substituição integral do trabalho humano, mas de reorganização de tarefas entre pessoas e sistemas, com diferentes graus de autonomia conforme o contexto. Atividades rotineiras podem ser delegadas, enquanto julgamento, relacionamento, estratégia e accountability permanecem predominantemente humanos.

Essa reorganização tende a exigir mudanças relevantes em job design, skills e modelos de gestão, porque muitas funções continuarão existindo, mas com uma distribuição completamente diferente entre execução, supervisão e tomada de decisão.

Organizações com workforce híbrida

Em determinados processos, AI Agents começarão a funcionar quase como uma nova categoria de recurso operacional, exigindo ownership, identidade, permissões, monitoring e avaliação de performance. Isso cria desafios completamente novos para Operating Model, governance e segurança.

Uma organização precisará saber quais agentes existem, quais sistemas podem acessar, quem responde por suas ações e quais controles impedem que ultrapassem os limites estabelecidos, aproximando a gestão de agentes de alguns conceitos tradicionalmente associados à gestão de identidades e workforce.

Interfaces cada vez mais naturais

Linguagem, voz e multimodalidade reduzem a necessidade de usuários conhecerem a estrutura específica dos sistemas com os quais interagem, fazendo com que intenção passe a ser uma interface cada vez mais importante. Essa mudança pode reduzir barreiras de utilização e ampliar acesso a capabilities tecnológicas mais sofisticadas.

Sistemas que anteriormente exigiam treinamento específico podem começar a ser utilizados através de linguagem natural, embora isso não elimine a necessidade de compreender corretamente o contexto e validar o resultado produzido.

Software cada vez mais dinâmico

Parte da lógica que anteriormente precisava ser rigidamente programada poderá ser interpretada em tempo de execução pelos modelos, alterando a própria maneira como aplicações são projetadas. Essa flexibilidade cria novas possibilidades, mas também exige mecanismos muito mais robustos de controle e observability.

Talvez a mudança mais profunda esteja justamente no fato de que AI começa a funcionar simultaneamente como tecnologia, interface e workforce digital, fazendo com que sua influência ultrapasse a discussão tradicional sobre mais uma categoria de software.

Principais desafios

Apesar das enormes oportunidades, GenAI continua apresentando desafios importantes, e alguns deles se tornaram ainda mais relevantes conforme a tecnologia ganhou maior capacidade de agir sobre sistemas e processos. O estágio atual exige que segurança, governance e operação evoluam no mesmo ritmo da funcionalidade.

Qualidade e confiabilidade

Modelos probabilísticos podem produzir respostas incorretas ou tomar caminhos inesperados, tornando evaluations, validation e monitoring componentes permanentes da arquitetura. Quanto mais crítico o processo, menor pode ser a tolerância a comportamentos imprevisíveis.

Por isso, qualidade precisa ser tratada de maneira sistemática, com métricas, testes, comparação entre versões e mecanismos de detecção de falhas adequados ao risco de cada caso de uso.

Privacidade e segurança

Aplicações podem lidar com informações confidenciais, propriedade intelectual, dados pessoais e sistemas críticos, exigindo controles claros sobre acesso, armazenamento e utilização. Esse desafio cresce conforme AI passa a circular por múltiplas fontes de informação dentro da empresa.

A questão deixa de ser apenas onde determinado prompt é processado e passa a envolver toda a cadeia de dados, ferramentas e sistemas acessados durante a execução.

Prompt Injection e tool abuse

Quando modelos passam a utilizar ferramentas e informações externas, novas superfícies de ataque aparecem e precisam ser consideradas desde o design. Um

sistema capaz de interpretar instruções em conteúdo não confiável pode ser manipulado de maneiras que não existiam em aplicações tradicionais.

Esse tipo de risco exige arquitetura específica, validação de ações e separação clara entre conteúdo, instruções e permissões.

Identidade de agentes

Um AI Agent capaz de executar operações precisa possuir identidade, privilégios mínimos e mecanismos claros de autorização, seguindo princípios equivalentes aos aplicados a usuários e workloads humanos. A organização precisa saber quem ou o que executou cada ação e sob qual autoridade.

Esse ponto será cada vez mais relevante conforme agentes deixarem ambientes experimentais e passarem a participar de processos operacionais críticos.

Governance em runtime

Policies escritas não são suficientes para sistemas que conseguem agir autonomamente, fazendo com que controles técnicos precisem impedir ações inadequadas no momento em que elas ocorreriam. Isso aproxima governance de enforcement e torna policy engines parte essencial da arquitetura.

Quanto maior o grau de autonomia, mais importante será transformar regras de governança em controles executáveis e auditáveis.

Viés e equidade

Sistemas treinados sobre grandes volumes de dados podem reproduzir ou amplificar padrões indesejáveis, principalmente quando utilizados para apoiar decisões relacionadas a pessoas. Esse risco exige avaliação contínua e compreensão do contexto em que cada modelo é aplicado.

A preocupação precisa ser proporcional ao impacto da decisão, já que um erro em geração de conteúdo possui consequências diferentes de um erro em crédito, contratação ou atendimento de populações vulneráveis.

Propriedade intelectual

Questões sobre dados de treinamento, conteúdos gerados e direitos autorais continuam evoluindo juridicamente e precisam ser acompanhadas conforme cada jurisdição. Empresas precisam entender claramente o regime de utilização associado aos modelos e aos outputs que produzem.

Esse tema se torna ainda mais importante quando AI é utilizada para produzir ativos comerciais, software, campanhas ou conteúdos destinados a terceiros.

Workforce e skills

AI modificará o desenho de funções e a distribuição de atividades, exigindo desenvolvimento de novas capacidades em vez de simplesmente disponibilização de ferramentas. A maior oportunidade provavelmente estará em profissionais que aprenderem a trabalhar em conjunto com sistemas inteligentes.

Capacitação precisa envolver não apenas domínio das ferramentas, mas também julgamento, validação, capacidade de formular problemas e entendimento de quando a resposta da tecnologia não deveria ser aceita.

Shadow AI

A facilidade de acesso a serviços públicos cria riscos semelhantes aos antigos movimentos de Shadow IT, mas potencialmente envolvendo dados muito mais sensíveis e decisões mais relevantes. Por isso, oferecer alternativas corporativas seguras pode ser tão importante quanto estabelecer restrições.

Quando a organização não oferece ferramentas adequadas, usuários naturalmente procuram alternativas externas, tornando governance muito mais difícil.

Segurança precisa acompanhar o aumento da autonomia

Creio que esse seja um dos pontos que mais mudou desde os primeiros debates sobre GenAI, porque enquanto a principal preocupação inicial estava relacionada ao vazamento de informações através de prompts ou à geração de conteúdo incorreto, sistemas agentic ampliam significativamente a superfície de risco ao permitir que modelos interajam diretamente com ferramentas.

Um agente capaz de apenas consultar dados possui um perfil de risco bastante diferente daquele autorizado a modificá-los, da mesma forma que um agente capaz de gerar uma sugestão de compra é muito diferente daquele que pode criar o pedido ou executar a transação. Essa diferença torna fundamental estabelecer níveis claros de autonomia e conectá-los ao tipo de processo, impacto potencial e capacidade de reversão.

Observe

O agente apenas acompanha informações e produz visibilidade, funcionando como uma camada adicional de percepção sem executar ações sobre o ambiente. Esse nível tende a possuir risco relativamente baixo e pode ser utilizado como ponto de entrada em diversos casos de uso.

Advise

O agente recomenda uma ação, mas uma pessoa continua responsável pela decisão, permitindo que a organização capture parte do benefício analítico sem transferir accountability. Esse modelo costuma ser especialmente adequado para processos nos quais julgamento permanece crítico.

Act with Approval

O agente prepara ou executa uma sequência, mas determinados passos exigem aprovação humana, combinando automação com controle explícito nos pontos de maior risco. Essa abordagem pode representar um equilíbrio bastante interessante entre produtividade e segurança.

Act Autonomously

O sistema possui autoridade para agir diretamente dentro de parâmetros definidos, algo que pode gerar enormes ganhos de velocidade e eficiência, mas exige controles muito mais sofisticados. Nem todo caso de uso deveria buscar esse estágio, e em muitos processos o melhor equilíbrio entre produtividade e risco pode permanecer permanentemente nos níveis intermediários.

GenAI e Data continuam inseparáveis

Existe também uma percepção que ficou ainda mais clara com a evolução da tecnologia: modelos extremamente poderosos não compensam dados empresariais ruins, porque uma organização pode contratar as melhores plataformas disponíveis e ainda assim obter resultados medíocres se suas informações forem inconsistentes, desatualizadas, fragmentadas ou inacessíveis.

GenAI cria até uma espécie de pressão positiva sobre Data Management porque torna problemas históricos muito mais visíveis, principalmente quando usuários passam a interagir diretamente com conhecimento corporativo através de linguagem natural. Quando um assistente fornece respostas erradas porque encontrou versões conflitantes de uma política, o verdadeiro problema talvez não esteja no modelo, mas no Knowledge Management e na qualidade das fontes disponíveis.

Da mesma maneira, quando um agente não consegue concluir determinado processo porque informações estão distribuídas em sistemas sem integração, o problema pode estar na arquitetura e não na capacidade de reasoning. AI frequentemente funciona como uma lupa sobre problemas que já existiam e obriga a organização a enfrentar questões de dados, integração e ownership que durante muitos anos puderam permanecer parcialmente escondidas.

Isso reforça a necessidade de conectar AI Strategy a Data Strategy, Enterprise Architecture, Cybersecurity e Operating Model, porque nenhuma dessas dimensões pode ser tratada isoladamente quando a tecnologia começa a participar diretamente da operação empresarial.

Medir valor é mais difícil do que medir adoção

Outro aprendizado importante está na diferença entre utilização e valor, pois é relativamente simples medir quantas pessoas possuem acesso a um Copilot, quantas interações foram realizadas ou quantos agentes estão ativos, mas muito mais difícil provar quanto valor empresarial foi realmente produzido.

Um usuário pode executar centenas de prompts sem alterar significativamente sua produtividade, enquanto outro pode utilizar a ferramenta poucas vezes para resolver atividades extremamente complexas e produzir impacto muito maior. Por isso, métricas de AI precisam evoluir para conectar utilização, qualidade, produtividade e business outcomes de maneira muito mais estruturada.

Adoption Metrics

Quantos usuários utilizam, com qual frequência e em quais funções continua sendo uma métrica relevante, principalmente para compreender se determinada capability está efetivamente sendo incorporada ao trabalho. Entretanto, adoção isoladamente não pode ser confundida com sucesso.

Productivity Metrics

Quanto tempo ou esforço foi reduzido ajuda a demonstrar impacto operacional, mas precisa ser analisado com cuidado para evitar que economia de tempo seja tratada automaticamente como valor. O que realmente importa é aquilo que a organização faz com a capacidade liberada.

Quality Metrics

Avaliar se o resultado produzido apresentou qualidade igual ou superior ao processo anterior é essencial, porque produtividade obtida à custa de maior retrabalho ou risco pode ser apenas aparente. AI precisa melhorar a relação entre velocidade e qualidade, e não simplesmente maximizar output.

Business Outcomes

Receita, conversão, satisfação, redução de risco, redução de custo ou outra métrica

diretamente ligada à estratégia representa o nível mais importante de avaliação. Quando a tecnologia consegue ser conectada a esse tipo de resultado, o Business Case se torna muito mais robusto.

Autonomy Metrics

A porcentagem do processo efetivamente concluída sem intervenção humana pode se tornar especialmente relevante na era agentic, principalmente quando combinada com métricas de qualidade e exceção. O objetivo não deveria ser maximizar autonomia indiscriminadamente, mas compreender em quais pontos ela efetivamente aumenta valor.

CIO Codex Framework – AI & ML

A integração de Artificial Intelligence & Machine Learning (AI & ML), incluindo subdomínios como Generative AI, na camada New Tech do CIO Codex Agenda Framework, representa uma das mais significativas revoluções tecnológicas na direção de um futuro automatizado e inteligente.

Este tema abraça o espectro completo da inteligência artificial, desde sistemas que automatizam tarefas rotineiras até algoritmos avançados capazes de gerar conteúdo e soluções inovadoras.

O conteúdo complementar detalha como AI & ML, em toda a sua extensão, podem ser efetivamente aplicadas para acelerar a transformação digital, impulsionando a inovação e garantindo uma vantagem competitiva robusta em todas as operações de negócio.

A introdução a AI & ML explora como essas tecnologias estão remodelando as estratégias de negócios e operações, permitindo uma nova era de automação e capacidades preditivas.

É discutida a aplicação de AI & ML na análise de dados complexos, no desenvolvimento de sistemas autônomos e na personalização de experiências do cliente.

Em particular, o foco é dado ao Generative AI, que representa a fronteira da criação de conteúdo e soluções inovadoras, oferecendo potencial para redefinir indústrias inteiras.

Este conteúdo aborda o desafio e a oportunidade de integrar AI & ML no tecido existente das operações de TI, desde a preparação e governança de dados até o desenvolvimento e implementação de modelos algorítmicos.

São analisadas as competências necessárias para construir equipes capazes de explorar o potencial da AI & ML, incluindo Generative AI, e as melhores práticas para gerenciar esses projetos complexos com responsabilidade e transparência.

São considerados os desafios operacionais e éticos de adotar AI & ML, enfatizando a importância da qualidade dos dados, da privacidade e da segurança.

São discutidas estratégias para a incorporação bem-sucedida dessas tecnologias avançadas e para a construção de uma cultura organizacional que suporte a inovação disruptiva e contínua.

Em conclusão, o conteúdo ressalta a necessidade de estabelecer métricas claras para avaliar o impacto de AI & ML, incluindo Generative AI, em termos de eficiência operacional, capacidade de inovação e contribuição para os resultados do negócio.

Visão prática

Ao considerarmos a implementação de tecnologias AI dentro das organizações, é crucial não apenas executar, mas sim desenvolver uma visão estratégica abrangente que aborde questões fundamentais.

Esta abordagem deve contemplar desde a identificação de processos, produtos e serviços afins, até a análise minuciosa dos casos de uso, modalidades de IA, investimentos necessários, e os riscos envolvidos.

A seguir são exploradas 5 questões essenciais quando do planejamento e elaboração de um roadmap para temas relacionados à AI:

1) – Identificação de Afinidades com a Tecnologia de IA

O primeiro passo crítico para a implementação bem-sucedida de Inteligência Artificial nas organizações envolve uma análise profunda para identificar quais processos, produtos ou serviços apresentam maior afinidade com essa tecnologia.

Este processo de avaliação começa com a compreensão de quais áreas da empresa são intensivas em dados e possuem operações repetitivas ou padrões previsíveis que podem ser otimizados por meio da IA.

Por exemplo, em uma instituição financeira, operações como análise de crédito podem ser significativamente aprimoradas utilizando modelos de aprendizado de máquina, que podem analisar grandes volumes de dados de crédito para identificar padrões e prever riscos de forma mais eficiente do que métodos tradicionais.

Outro exemplo pode ser encontrado no setor de atendimento ao cliente, onde chatbots alimentados por IA podem gerenciar consultas de rotina, liberando funcionários

humanos para lidar com casos mais complexos.

Além de identificar onde a IA pode ser aplicada, é crucial avaliar a maturidade atual dos processos tecnológicos da organização.

A existência de uma infraestrutura de dados robusta e uma cultura organizacional que apoia a inovação digital são pré-requisitos para que a implementação de soluções de IA seja bem-sucedida. Assim, o diagnóstico deve também focar na prontidão tecnológica e na disposição cultural para adotar novas soluções.

2) – Escolha da Modalidade de IA para cada Caso de Uso

Uma vez identificados os processos e áreas com potencial para a aplicação de IA, a próxima etapa é determinar qual modalidade de IA se adapta melhor a cada caso de uso específico.

A decisão deve considerar o objetivo do projeto de IA, os tipos de dados disponíveis e os resultados esperados.

Por exemplo, se o objetivo é melhorar a interação com o cliente através do entendimento e resposta a suas necessidades em tempo real, o processamento de linguagem natural (NLP) pode ser a modalidade mais adequada.

O NLP permite que sistemas computacionais compreendam, interpretem e respondam a textos humanos de maneira eficaz, facilitando uma comunicação mais natural e intuitiva com os usuários.

Em contrapartida, se a organização busca otimizar suas operações logísticas, modelos preditivos de aprendizado de máquina podem ser implementados para prever demandas de estoque e otimizar rotas de entrega.

Esses modelos são capazes de analisar históricos de dados complexos e identificar tendências e padrões que humanos poderiam não perceber.

A escolha da modalidade de IA também deve levar em consideração as limitações técnicas, como a qualidade e quantidade dos dados disponíveis.

Modelos de aprendizado profundo, por exemplo, requerem grandes volumes de dados de alta qualidade para treinamento, o que pode ser um desafio em ambientes com dados limitados ou de baixa qualidade.

3) – Análise de Business Case: Investimentos Versus Retornos

Para cada potencial aplicação de Inteligência Artificial, a criação de um business case detalhado é essencial.

Este documento deve avaliar minuciosamente os custos e benefícios associados, tanto

de curto quanto de longo prazo.

É crucial que cada caso de uso de IA seja justificado não só em termos de benefícios diretos, como eficiência operacional e aumento de receita, mas também considerando benefícios indiretos, como melhorias na satisfação do cliente e fortalecimento da imagem da marca.

Por exemplo, a implementação de um sistema de IA para personalização de ofertas para clientes pode requerer investimentos iniciais significativos em tecnologia e treinamento de equipe, mas os retornos podem incluir um aumento notável na fidelização de clientes e no valor médio de compra.

A análise deve também estimar o tempo necessário para que os investimentos se paguem (payback) e o retorno sobre o investimento (ROI) projetado para os próximos anos.

Neste contexto, é importante incorporar variáveis como a velocidade de adoção da tecnologia pelos usuários, a escalabilidade das soluções e potenciais custos ocultos, como manutenção e atualizações tecnológicas necessárias para sustentar a iniciativa ao longo do tempo.

Modelos financeiros, como análise de fluxo de caixa descontado, podem ser utilizados para estimar o valor presente líquido (VPL) e a taxa interna de retorno (TIR), proporcionando uma base sólida para a tomada de decisão.

4) – Investimentos “Reais” para Implementação e Manutenção

Implementar tecnologias de IA vai além da simples aquisição de software ou hardware; envolve uma série de investimentos que podem ser substanciais.

Primeiramente, muitas soluções de IA requerem subscrições de serviços SaaS que podem ter custos recorrentes significativos.

Além disso, a contratação e a formação de equipes especializadas são essenciais, pois a gestão e operação de sistemas de IA requerem habilidades específicas que muitas vezes não estão presentes internamente nas organizações.

Outro aspecto importante é a adequação da infraestrutura de TI existente.

A implementação de IA frequentemente exige atualizações significativas em hardware e software para suportar o processamento intensivo de dados. Isso pode incluir, por exemplo, a expansão de capacidades de armazenamento de dados ou a atualização de sistemas de segurança para proteger os dados manipulados.

A integração de sistemas de IA com sistemas legados também representa um desafio técnico e financeiro.

Muitas vezes, sistemas mais antigos não são projetados para interagir com tecnologias baseadas em IA requerendo adaptações ou até mesmo a substituição de sistemas existentes, o que pode elevar significativamente os custos de projeto.

Finalmente, não se pode ignorar os custos contínuos associados à manutenção e atualização dos sistemas de IA.

Estes sistemas precisam ser constantemente treinados com novos dados para manter sua eficácia, e as soluções de software precisam ser atualizadas para se adaptar a novas ameaças de segurança e mudanças na legislação, especialmente no que diz respeito à privacidade de dados.

5) – Avaliação dos Riscos de Adoção Versus Não Adoção

A decisão de implementar tecnologias de AI em uma organização envolve não apenas a análise de benefícios potenciais, mas também uma avaliação cuidadosa dos riscos associados.

Esses riscos podem ser divididos em dois grandes grupos: os riscos de prosseguir com a iniciativa de IA e os riscos de optar por não a adotar.

Riscos de Adoção da IA

• Investimento Inicial Elevado Sem Garantias de Retorno: A implementação de soluções de IA frequentemente exige investimentos substanciais em tecnologia, treinamento e reestruturação de processos. Existe o risco de que esses investimentos não se traduzam em melhorias de desempenho ou ganhos financeiros dentro do prazo esperado, especialmente se a implementação não for bem planejada ou se a tecnologia escolhida não se adequar às necessidades da empresa.

• Complexidade Técnica e Falhas Potenciais: Sistemas de IA são complexos e podem falhar de maneiras inesperadas, especialmente se forem mal configurados ou se operarem em ambientes variáveis. Erros de IA podem levar a decisões empresariais equivocadas, interrupções operacionais ou problemas de segurança.

• Dependência de Fornecedores e Tecnologia: Ao adotar soluções de IA, as organizações muitas vezes se tornam dependentes dos fornecedores dessas tecnologias para suporte contínuo, atualizações e manutenção. Isso pode limitar a flexibilidade operacional e aumentar os custos a longo prazo.

• Questões Éticas e de Conformidade: A implementação de IA envolve desafios significativos em termos de ética e conformidade legal, especialmente relacionados à privacidade de dados e ao viés algorítmico. Falhas em abordar adequadamente essas questões podem resultar em danos reputacionais e penalidades regulatórias.

Riscos de Não Adotar a IA

• Perda de Competitividade: À medida que mais organizações adotam IA para otimizar operações, personalizar serviços e inovar em produtos, as empresas que optam por não adotar essas tecnologias podem se encontrar em desvantagem competitiva. A incapacidade de oferecer serviços comparáveis ou de operar com a mesma eficiência pode resultar em perda de mercado e de clientes.

• Obsolescência Tecnológica: A tecnologia evolui rapidamente, e sistemas que não incorporam IA podem rapidamente se tornar obsoletos. A falta de atualização tecnológica pode levar a ineficiências operacionais e aumentar os custos de manutenção de sistemas legados.

• Incapacidade de Atender às Expectativas do Cliente: Os consumidores estão cada vez mais esperando experiências personalizadas e eficientes que frequentemente só podem ser fornecidas através de tecnologias avançadas como a IA. A não adoção pode resultar em uma percepção de marca antiquada e em uma base de clientes insatisfeita.

Portanto, a decisão de adotar ou não a IA deve ser baseada em uma compreensão clara dos riscos e benefícios potenciais.

É vital que as organizações não apenas considerem os custos e desafios técnicos, mas também avaliem como a adoção, ou a falta dela, alinha-se com suas estratégias de longo prazo e objetivos de mercado.

A análise de risco deve ser um processo contínuo, adaptando-se às mudanças no ambiente de negócios e na tecnologia para garantir que a organização permaneça resiliente e competitiva.

Evolução Cronológica

A trajetória da Inteligência Artificial (IA) e do Machine Learning (ML) é marcada por desenvolvimentos significativos que refletem as mudanças nas demandas tecnológicas e empresariais.

A seguir é apresentada uma visão detalhada da evolução cronológica da IA e ML, desde suas origens conceituais até as inovações mais recentes, ilustrando como essas tecnologias revolucionaram a infraestrutura de TI nas organizações.

A IA e o ML continuam a evoluir, respondendo tanto às oportunidades tecnológicas quanto aos desafios operacionais.

À medida que novas tecnologias emergem e os custos de infraestrutura flutuam, as estratégias de TI devem permanecer ágeis e adaptativas.

A capacidade de uma organização de se adaptar eficientemente será crucial para manter a competitividade e a inovação em um ambiente empresarial que é, por natureza, volátil e em constante evolução.

1) – A Gênese da Inteligência Artificial (Anos 1950 – 1980)

• Origens Conceituais: Nos anos 1950, pesquisadores como Alan Turing e John McCarthy introduziram os fundamentos da IA. Turing propôs o “Teste de Turing” como um critério para a inteligência de máquinas, enquanto McCarthy cunhou o termo “Inteligência Artificial” e organizou a famosa conferência de Dartmouth em 1956, que é considerada o ponto de partida formal do campo da IA.

• Primeiros Avanços: Durante as décadas de 1960 e 1970, os primeiros programas de IA, como o ELIZA e o SHRDLU, demonstraram capacidades básicas de processamento de linguagem natural e resolução de problemas. A pesquisa em redes neurais artificiais também começou a ganhar tração, embora com limitações significativas devido à falta de poder computacional e dados.

2) – O Inverno da IA e o Ressurgimento (Anos 1980 – 2000)

• Inverno da IA: O entusiasmo inicial foi seguido por um período de estagnação conhecido como o “Inverno da IA”, onde a falta de resultados práticos e a limitação tecnológica resultaram em cortes de financiamento e interesse. Durante os anos 1980, apesar dos desafios, a pesquisa continuou, com avanços em sistemas especialistas, que utilizavam regras baseadas em conhecimento para resolver problemas específicos.

• Ressurgimento: Nos anos 1990, a IA começou a ressurgir com o advento de computadores mais poderosos e a explosão da Internet. Os algoritmos de machine learning, particularmente o aprendizado supervisionado e não supervisionado, começaram a mostrar resultados promissores em áreas como reconhecimento de padrões e análise de dados.

3) – A Era do Big Data e Machine Learning (Anos 2000 – 2010)

• Explosão de Dados: A disponibilidade massiva de dados e o avanço das tecnologias de armazenamento e processamento criaram um terreno fértil para o machine learning. Algoritmos como árvores de decisão, regressão logística e redes neurais se tornaram comuns em aplicações de negócios e pesquisa.

• Deep Learning: O início dos anos 2010 viu o surgimento do deep learning, uma subárea do machine learning baseada em redes neurais profundas. Modelos como o AlexNet, desenvolvido por Geoffrey Hinton e sua equipe, revolucionaram a área de visão computacional ao vencer competições de reconhecimento de imagem, estabelecendo um novo padrão para a precisão e capacidade de generalização.

4) – A Era da Inteligência Artificial Pervasiva (2010 – Presente)

• IA na Vida Cotidiana: A partir de 2010, a IA e o ML começaram a penetrar em quase todos os aspectos da vida cotidiana e empresarial. Aplicações como assistentes virtuais (Siri, Alexa), veículos autônomos, diagnósticos médicos assistidos por IA, e sistemas de recomendação em plataformas de streaming e e-commerce se tornaram comuns.

• IA e Ética: Com a crescente influência da IA, surgiram preocupações éticas e de governança. Questões como viés algorítmico, privacidade de dados e o impacto no mercado de trabalho passaram a ser temas centrais nos debates sobre a adoção de IA.

• Futuro da IA: As tendências atuais incluem a integração de IA com outras tecnologias emergentes, como Internet das Coisas (IoT) e Computação em Borda (Edge Computing), para criar sistemas ainda mais responsivos e inteligentes. O desenvolvimento contínuo em áreas como processamento de linguagem natural, aprendizado por reforço e IA explicável promete ampliar ainda mais as fronteiras do que é possível com a inteligência artificial.

Em suma, a evolução da IA e ML tem sido uma jornada de altos e baixos, marcada por avanços tecnológicos significativos e desafios complexos.

À medida que essas tecnologias continuam a se desenvolver, elas prometem transformar ainda mais a forma como vivemos e trabalhamos, exigindo uma abordagem cuidadosa e ética para sua implementação e uso.

Concluindo

Como exploro em diferentes artigos e apresentações, integrar GenAI aos negócios deixou definitivamente de ser apenas uma questão tecnológica e passou a representar uma discussão sobre a própria forma como as empresas operam, tomam decisões, acessam conhecimento e distribuem trabalho entre pessoas e sistemas digitais. Os líderes de tecnologia precisam se posicionar não apenas como executores responsáveis por selecionar modelos ou plataformas, mas como estrategistas capazes de orientar a organização sobre onde AI efetivamente produz valor, quais processos precisam ser redesenhados, qual nível de autonomia faz sentido e quais riscos precisam permanecer sob controle.

Talvez uma das maiores mudanças desde o início da GenAI esteja justamente no fato de que o debate já não precisa provar se a tecnologia possui capacidade real, pois essa etapa foi amplamente superada. O verdadeiro desafio agora está em separar capacidade tecnológica de valor empresarial, porque um modelo pode ser impressionante e ainda assim não justificar determinado projeto, um agente pode

automatizar dezenas de etapas e mesmo assim gerar economics inadequada e um Copilot pode ser utilizado diariamente sem produzir impacto mensurável suficientemente relevante.

Essa distinção deveria orientar qualquer estratégia madura de AI, principalmente porque a entrada definitiva na era dos AI Agents aumenta significativamente a importância da governança ao transformar risco potencial em risco operacional. Enquanto uma resposta incorreta pode ser revisada por uma pessoa antes de qualquer consequência, uma ação autônoma eventualmente acontece antes que alguém tenha oportunidade de intervir, fazendo com que controles de identidade, autorização, observability, policy enforcement e human oversight precisem evoluir juntamente com a tecnologia.

A mesma lógica vale para segurança e privacidade, pois não considero realista imaginar que empresas possam esperar todos os desafios serem completamente resolvidos antes de adotar AI, já que esse momento provavelmente nunca chegará. Tecnologias complexas sempre carregam riscos, e a questão está em desenvolver estruturas capazes de entender, controlar e aceitar conscientemente esses riscos de acordo com o valor potencial de cada caso de uso.

Outro aspecto que considero essencial é evitar transformar a corrida por AI em uma corrida por quantidade de projetos, porque possuir cem Proofs of Concept não significa necessariamente estar mais avançado do que uma empresa que possui dez casos de uso efetivamente industrializados, integrados aos processos e produzindo resultados concretos. A maturidade deveria ser medida por valor, escala, segurança e sustentabilidade, e não simplesmente pela quantidade de iniciativas anunciadas ou ferramentas disponibilizadas.

Também devemos evitar repetir alguns erros observados em outras grandes ondas tecnológicas, porque Cloud ensinou que migração sem transformação pode simplesmente transferir complexidade de lugar, Agile mostrou que adotar cerimônias sem mudar mentalidade não cria agilidade e Digital Transformation demonstrou que adquirir plataformas não transforma automaticamente uma organização. AI provavelmente ensinará exatamente a mesma lição, pois a tecnologia é habilitadora, enquanto transformação continua dependendo de pessoas, processos, dados, arquitetura, cultura e liderança.

Por isso, considero especialmente importante que cada organização estabeleça um portfolio de casos de uso e aprenda a tratá-los de maneira distinta, porque alguns são excelentes candidatos para produtividade individual, outros justificam automação completa, determinados processos exigem human-in-the-loop e alguns talvez simplesmente não devam utilizar GenAI. Não existe qualquer razão para imaginar que

uma única abordagem seja adequada para toda a empresa, e essa capacidade de escolher talvez seja mais importante do que a própria capacidade de implementar.

Existe ainda a dimensão do risco de não fazer nada, porque à medida que AI se torna parte do custo operacional, do desenvolvimento de software, do atendimento ao cliente e da produtividade de diferentes funções, empresas que simplesmente ignorarem sua utilização poderão enfrentar uma diferença crescente de performance em relação a competidores capazes de incorporá-la de maneira efetiva. Esse risco, entretanto, não justifica adoção irresponsável, mas reforça a necessidade de encontrar um equilíbrio entre velocidade, aprendizado e controle.

Creio que a resposta correta está justamente em mover-se suficientemente rápido para aprender, governar suficientemente bem para controlar, medir suficientemente bem para saber se existe valor e preservar flexibilidade suficiente para acompanhar uma tecnologia que continuará mudando em enorme velocidade. Essa combinação provavelmente será muito mais importante do que qualquer decisão específica sobre qual modelo ou fornecedor escolher em determinado momento.

A IA generativa já ultrapassou o estágio em que poderia ser vista apenas como ferramenta para criação de conteúdo e começa a funcionar simultaneamente como camada de interação com conhecimento, capability de automação, interface para software e, progressivamente, uma nova forma de workforce digital. Isso amplia extraordinariamente seu potencial, mas também aumenta proporcionalmente nossa responsabilidade sobre como utilizá-la.

Para CIOs e líderes de tecnologia, talvez esse seja o verdadeiro desafio da próxima etapa, porque não se trata mais de convencer a organização de que AI é importante, mas de transformar essa importância em uma estratégia empresarial concreta, escolher os casos de uso certos, industrializar aquilo que funciona, abandonar aquilo que não produz valor e criar um modelo de governança capaz de evoluir juntamente com a tecnologia.

Creio que as empresas que conseguirem fazer isso não serão necessariamente aquelas que utilizarem mais AI, mas aquelas que conseguirem utilizar AI exatamente nos pontos em que ela modifica de maneira relevante a economics, a experiência ou a capacidade de execução do negócio. E talvez seja justamente aí que a pergunta inicial encontre sua melhor resposta, pois os principais usos para GenAI não são uma lista fixa de aplicações, mas todos os pontos nos quais conhecimento, linguagem, decisão e execução podem ser reorganizados de maneira suficientemente melhor para justificar uma nova forma de trabalhar.