Nesses tempos de explosão exponencial do uso e divulgação da Inteligência Artificial, com destaque para a Generative AI, não faltam artigos, estudos, webinars, eventos e matérias buscando equalizar o conhecimento geral sobre o tema, algo absolutamente compreensível quando consideramos a velocidade com que novas capacidades surgem e passam quase imediatamente dos laboratórios de pesquisa para produtos que podem ser utilizados por milhões de pessoas e empresas.

E aqui fica a recomendação de um webinar do Gartner, com uma visão muito bem estruturada sobre o tema e que ajuda a organizar alguns dos conceitos fundamentais relacionados à GenAI e à forma como essa tecnologia pode ser encarada pelas organizações:

https://webinar.gartner.com/517294/agenda/session/1178143

No cenário atual, marcado por rápidas transformações tecnológicas e pela progressiva digitalização de praticamente todos os setores da economia, a Inteligência Artificial assume um papel central como motor de inovação, produtividade e transformação, deixando rapidamente de ser apenas uma tecnologia exploratória para se tornar uma capability que começa a ser incorporada de forma transversal aos produtos, processos, plataformas e modelos operacionais das empresas.

A importância dessa transformação não decorre apenas da capacidade de modelos generativos produzirem textos, imagens, códigos, vídeos ou análises cada vez mais sofisticadas, mas principalmente da combinação entre essas capacidades e o acesso aos dados, aplicações e processos corporativos, algo que amplia enormemente o potencial da tecnologia e, ao mesmo tempo, aumenta significativamente a complexidade de sua adoção no mundo enterprise.

Hoje em dia se mostra essencial explorar esses novos conceitos e tecnologias buscando uma reflexão mais profunda sobre como líderes e organizações podem se posicionar estrategicamente para aproveitar as capacidades da Inteligência Artificial, sem cair na armadilha relativamente comum de assumir que uma tecnologia extremamente poderosa necessariamente precisa ser utilizada em qualquer situação apenas porque o mercado está concentrando enorme atenção sobre ela.

E vale reiterar mais uma vez: considere que isso é só o começo, pois apesar da enorme evolução já observada, a tecnologia continua avançando em uma velocidade impressionante tanto em profundidade, com modelos cada vez mais capazes e eficientes, quanto em abrangência, com novas aplicações surgindo continuamente e expandindo as fronteiras daquilo que pode ser automatizado, acelerado ou completamente redesenhado.

Não imaginamos exatamente aonde vamos chegar em alguns anos, seja pela evolução das capacidades cognitivas dos modelos, seja pela redução de custos, seja pela criação de novas arquiteturas ou pela integração progressiva da AI com praticamente todas as categorias de software empresarial, mas me parece razoável assumir que a tecnologia ainda está distante de atingir um estado de maturidade ou estabilidade suficiente para permitir que alguém consiga antecipar claramente seus limites.

De qualquer forma, vale conhecer a abordagem apresentada pelo Gartner e utilizá-la como referência para refletir sobre como estruturar a adoção dentro da própria empresa, buscando fazer uso da tecnologia da forma que efetivamente gere valor e, ao mesmo tempo, reduzindo o risco de seguir por caminhos que pareçam interessantes durante a fase inicial de experimentação, mas que posteriormente se mostrem difíceis

de escalar, operar, governar ou justificar economicamente.

Creio que uma parte relevante do desafio empresarial relacionado à Inteligência Artificial deixou de estar apenas na capacidade tecnológica dos modelos e passou a se concentrar em como aparar as arestas relacionadas a segurança, privacidade, compliance, governança, arquitetura, dados, custos, integração, monitoramento e todos os demais elementos necessários para transformar uma ferramenta tecnicamente impressionante em uma capability confiável dentro do mundo enterprise.

O webinar do Gartner

O webinar detalha diversos aspectos da Inteligência Artificial, especialmente focando na Generative AI, suas definições, aplicações e as precauções necessárias ao adotá-la em ambientes corporativos, apresentando a tecnologia não apenas como uma ferramenta capaz de produzir novos conteúdos a partir de grandes volumes de dados existentes, mas como um elemento transformador que exige uma nova abordagem de engajamento técnico e estratégico por parte das lideranças empresariais.

Essa diferenciação é bastante importante porque a GenAI não deveria ser compreendida simplesmente como uma evolução de ferramentas tradicionais de automação, uma vez que modelos generativos possuem características probabilísticas e capacidades de interpretação que alteram substancialmente a relação entre o usuário, a informação e o próprio software.

O webinar também destaca que muitas tecnologias classificadas genericamente como Artificial Intelligence possuem raízes em conceitos estatísticos, matemáticos e computacionais desenvolvidos há décadas, mas que passaram por uma transformação significativa com o aumento da disponibilidade de dados, evolução da capacidade de processamento, novas arquiteturas de modelos e disponibilidade de infraestrutura de Cloud em escala global.

Essa combinação produziu uma mudança estrutural importante, pois capabilities que anteriormente dependiam de equipes altamente especializadas passaram a ser disponibilizadas diretamente aos usuários através de interfaces extremamente simples baseadas em linguagem natural, fazendo com que uma parcela muito maior das pessoas pudesse experimentar diretamente os benefícios e limitações da Inteligência Artificial.

Modelos como os GPTs, baseados na arquitetura Transformer, são exemplos claros dessa evolução e demonstraram capacidade para geração de texto, tradução, sumarização, programação, classificação, interpretação de conteúdo e inúmeras outras atividades, embora permaneçam sujeitos a limitações relevantes relacionadas a

precisão, confiabilidade, vieses, hallucinations e dificuldade de explicar exatamente como determinadas respostas foram produzidas.

Importante notar é a ênfase dada às aplicações práticas da GenAI em diversas áreas, desde atendimento ao cliente até Software Development, passando por marketing, Analytics, gestão do conhecimento, análise documental e inúmeras outras atividades nas quais a capacidade de compreender e gerar informação pode aumentar significativamente a produtividade humana.

Ao mesmo tempo, quanto mais essas tecnologias deixam de ser utilizadas apenas para auxiliar uma pessoa e passam a ser integradas diretamente aos processos empresariais, maior se torna a necessidade de mecanismos capazes de assegurar que os resultados produzidos estejam dentro dos níveis aceitáveis de qualidade, segurança, compliance e risco.

De Generative AI para Agentic AI

Uma das evoluções mais relevantes do mercado está justamente na passagem progressiva de uma AI predominantemente generativa para modelos capazes de atuar de maneira mais autônoma, consolidando o conceito de Agentic AI e ampliando consideravelmente o impacto potencial dessas tecnologias sobre os processos empresariais.

Em um modelo tradicional de utilização da GenAI, uma pessoa fornece uma pergunta ou instrução, recebe uma resposta e posteriormente decide o que fazer com aquela informação, mantendo o usuário como elemento responsável por praticamente toda a execução posterior.

Com Agentic AI, essa fronteira começa a mudar, pois um agente pode receber determinado objetivo, interpretar o contexto existente, decompor a atividade em etapas, consultar informações, utilizar diferentes ferramentas, acessar aplicações, executar determinadas ações e avaliar se o resultado obtido atende ao objetivo originalmente definido.

Essa mudança pode parecer apenas uma evolução natural da tecnologia, mas suas implicações são profundas, pois existe uma diferença significativa entre um sistema que fornece uma recomendação para uma pessoa e outro que recebe autorização para efetivamente atuar sobre determinado processo empresarial.

Um agente que apenas resume documentos possui um perfil de risco relativamente limitado, enquanto um agente autorizado a alterar registros, emitir comunicações, modificar configurações, realizar transações ou acionar outros sistemas passa a

ocupar uma posição completamente diferente dentro da arquitetura corporativa.

Por essa razão, uma das discussões mais atuais do Gartner se concentra justamente na necessidade de diferenciar AI Agents de acordo com seus respectivos níveis de autonomia, evitando uma abordagem simplista na qual todos sejam governados da mesma maneira independentemente de sua capacidade de ação ou do nível de acesso concedido.

Essa lógica parece absolutamente coerente, pois um agente com acesso exclusivamente read-only deveria estar sujeito a controles diferentes daqueles exigidos por um agente capaz de efetuar alterações ou agir autonomamente, fazendo com que governança e autonomia precisem evoluir de maneira proporcional.

E esse talvez seja um dos pontos que melhor ilustra a mudança de estágio da tecnologia, pois durante algum tempo o principal debate estava relacionado à qualidade das respostas produzidas pela AI, enquanto agora passa a ser igualmente necessário discutir identidade, autorização, accountability, rastreabilidade, limites de autonomia e mecanismos capazes de interromper comportamentos inadequados.

Reflexões sobre a implementação de GenAI

A adoção da GenAI representa um divisor de águas no panorama tecnológico atual, sobretudo porque combina uma enorme capacidade de inovação com uma barreira inicial de entrada relativamente baixa, permitindo que praticamente qualquer organização realize experimentos e Proofs of Concept antes mesmo de possuir uma estratégia empresarial suficientemente estruturada sobre o tema.

Essa facilidade pode ser extremamente positiva durante uma fase exploratória, mas também pode produzir a falsa impressão de que escalar AI será simplesmente uma continuação natural daqueles primeiros experimentos, quando na realidade existe uma diferença significativa entre construir uma demonstração funcional e operar uma capability empresarial crítica de forma segura, resiliente, economicamente previsível e aderente aos padrões corporativos.

Abaixo exploro algumas reflexões estratégicas que considero especialmente importantes sobre o impacto e a implementação da GenAI nas organizações, atualizando o debate para um cenário no qual Generative AI, AI Agents e aplicações empresariais começam progressivamente a convergir.

Generative AI como Serviço Cloud SaaS

A tendência atual direciona uma parcela significativa da Inteligência Artificial Generativa para modelos de consumo baseados em Cloud, algo bastante natural quando consideramos o enorme volume de infraestrutura, processamento, energia, desenvolvimento e atualização necessário para criar e operar Foundation Models competitivos.

Esse movimento permite escalabilidade e acessibilidade sem precedentes, reduzindo drasticamente a barreira de entrada para organizações que desejam explorar capacidades avançadas de AI sem precisar realizar investimentos bilionários em infraestrutura e pesquisa, algo que somente um grupo relativamente pequeno de empresas conseguiria economicamente justificar.

Empresas muito grandes podem eventualmente desenvolver modelos próprios, customizar modelos existentes ou operar modelos open-weight em ambientes privados, principalmente quando existirem requisitos específicos relacionados a soberania, segurança, propriedade intelectual ou performance, mas a grande maioria do mercado provavelmente continuará consumindo essas capabilities através das grandes plataformas tecnológicas.

Entretanto, acredito que a discussão atual exige uma nuance adicional, pois o cenário deixou de ser simplesmente escolher entre consumir um SaaS ou construir alguma coisa internamente, surgindo uma variedade crescente de alternativas como APIs, Model as a Service, private endpoints, modelos open-weight, Small Language Models, plataformas de agentes, ambientes dedicados e diferentes combinações entre Cloud e infraestrutura própria.

O ponto estratégico passa então a ser muito menos decidir se AI será consumida através da Cloud e muito mais definir qual arquitetura oferece a melhor combinação entre capability, custo, segurança, governança, portabilidade, performance e integração para cada categoria de utilização.

Essa preocupação se torna ainda maior quando diferentes aplicações empresariais passam a incorporar seus próprios modelos e agentes, criando a possibilidade de que uma mesma organização consuma simultaneamente serviços de AI provenientes de dezenas de fornecedores diferentes.

Impacto econômico e estratégico para vendors

de tecnologia

O ponto acima fortalece diretamente a posição dos grandes fornecedores de tecnologia, como Amazon AWS, Microsoft Azure, Google Cloud e inúmeros outros players que passaram a investir volumes extraordinários de recursos na construção de modelos, infraestrutura, plataformas e serviços relacionados à Artificial Intelligence.

Essa tendência não apenas representa uma nova fonte de receita para os fornecedores, mas cria uma oportunidade estratégica muito mais ampla, pois AI começa a ser incorporada em praticamente todas as camadas do technology stack e deixa de existir apenas como uma categoria isolada de produto.

ERP, CRM, HCM, Productivity Tools, Development Platforms, Databases, Cybersecurity, Analytics, Customer Service e praticamente qualquer outra categoria relevante do enterprise software começa a incorporar algum tipo de AI, fazendo com que a tecnologia passe progressivamente a fazer parte natural das soluções já consumidas pelas organizações.

Nesse cenário, talvez o desafio futuro não seja convencer as empresas a adotar Inteligência Artificial, mas administrar a quantidade de funcionalidades baseadas em AI que será automaticamente introduzida pelos próprios fornecedores das soluções existentes.

Isso pode criar uma situação semelhante à expansão inicial das aplicações SaaS, quando a facilidade de contratação permitiu que diferentes áreas acumulassem soluções redundantes ou pouco integradas antes que arquitetura e governança conseguissem estruturar adequadamente aquele crescimento.

Portanto, a adoção de AI provavelmente exigirá uma visão de portfolio muito mais ampla, na qual não seja suficiente analisar individualmente cada novo recurso oferecido por determinado fornecedor, sendo necessário compreender como o conjunto dessas capabilities se integra à arquitetura corporativa e à estratégia tecnológica de longo prazo.

GenAI como potencial “evento de extinção”

A GenAI tem potencial para funcionar como uma espécie de “evento de extinção” para empresas que falham completamente em compreender ou incorporar seus impactos, principalmente porque uma tecnologia capaz de alterar simultaneamente produtividade, velocidade, qualidade, personalização e custo pode produzir diferenças competitivas acumulativas ao longo do tempo.

Não significa necessariamente que qualquer empresa que não adotar rapidamente

toda novidade associada à AI esteja condenada, mas existe um risco real de que organizações mais eficientes passem a operar com estruturas de custo, velocidade de decisão e capacidade de inovação significativamente melhores do que concorrentes que permanecerem excessivamente presos aos modelos tradicionais.

Essa diferença pode inicialmente parecer pequena, mas tende a se acumular, pois uma empresa capaz de desenvolver software mais rapidamente, reduzir determinadas atividades operacionais, melhorar a personalização, acelerar análises e ampliar a produtividade do knowledge worker pode gradualmente construir uma vantagem que se manifesta em diversos pontos da cadeia de valor simultaneamente.

Nesse sentido, é possível que não exista uma segunda oportunidade simples para empresas que esperem tempo demais antes de entender as mudanças estruturais trazidas pela tecnologia, principalmente em setores nos quais novos competidores AI-native possam surgir com modelos operacionais significativamente diferentes daqueles das organizações tradicionais.

Entretanto, vejo hoje um risco tão importante quanto a própria não adoção: adotar AI de maneira desordenada e confundir experimentação com transformação.

Uma organização pode possuir centenas de Proofs of Concept, milhares de usuários utilizando diferentes ferramentas e uma quantidade crescente de iniciativas relacionadas à AI sem que isso signifique que uma capability empresarial tenha efetivamente sido construída.

Shadow AI, duplicação de soluções, ausência de standards, dependência excessiva de fornecedores, custos imprevisíveis, exposição de dados e falta de clareza sobre accountability podem transformar uma agenda aparentemente inovadora em uma importante fonte de Technical Debt e risco operacional.

Por isso, talvez a verdadeira distinção futura não esteja entre empresas que “usam AI” e empresas que “não usam AI”, mas entre organizações capazes de industrializar a tecnologia de forma sustentável e aquelas que permanecerão eternamente presas a uma sucessão de experimentos.

Custos e planejamento estratégico na adoção de GenAI

A adoção de GenAI exige um planejamento financeiro e estratégico cuidadoso e é melhor que as empresas tenham isso bastante claro durante a criação e aprovação de seus Business Plans e Investment Plans, pois o custo aparente de uma experimentação inicial pode ser muito diferente daquele necessário para sustentar uma utilização

empresarial em escala.

É relativamente fácil criar um pequeno piloto utilizando uma API ou uma ferramenta SaaS e demonstrar resultados interessantes para um grupo limitado de usuários, mas colocar a mesma solução em produção pode exigir investimentos relevantes em arquitetura, integração, segurança, dados, Observability, controles, suporte, governança e capacitação.

Os custos associados à implementação e escalabilidade das soluções de GenAI podem abranger infraestrutura tecnológica, armazenamento, processamento, consumo de tokens, inference, contratação de talentos, treinamento, ferramentas especializadas e inúmeros componentes que não necessariamente aparecem de forma evidente no primeiro Business Case.

Além disso, aplicações agentic adicionam uma variável importante a essa equação, pois uma única atividade solicitada por um usuário pode gerar dezenas ou centenas de interações entre modelos, ferramentas, APIs e sistemas antes de produzir um resultado final.

Isso significa que o custo de AI precisa ser progressivamente analisado não apenas por usuário ou licença, mas também no nível da transação e do processo empresarial.

Quanto custa um agente para concluir determinada atividade?

Quantas interações com modelos são necessárias?

Qual percentual das atividades termina com sucesso?

Quantas exigem intervenção humana?

Qual o custo de uma execução malsucedida?

Qual benefício foi efetivamente produzido?

Essas perguntas tendem a ganhar importância conforme AI evolui da experimentação para a industrialização e o mercado começa a exigir a mesma disciplina econômica aplicada a qualquer outra capability corporativa.

Talvez vejamos surgir uma disciplina cada vez mais próxima do que FinOps representou para Cloud Computing, voltada para acompanhar o consumo, otimizar modelos, controlar recursos e assegurar que a economia das aplicações de AI permaneça sustentável.

Lições da implementação da tecnologia Cloud

É vital que as empresas apliquem as lições aprendidas com a adoção da tecnologia

Cloud ao implementar soluções de GenAI, pois algumas das experiências mais importantes da transformação para Cloud não estavam necessariamente relacionadas à tecnologia em si, mas à forma como organizações interpretaram inicialmente aquilo que significava adotar uma nova plataforma.

Estratégias precipitadas como o conhecido “lift and shift” permitiram migrar rapidamente workloads tradicionais, mas nem sempre conseguiram capturar todo o valor da tecnologia Cloud, fazendo com que algumas empresas descobrissem posteriormente que haviam basicamente mudado a localização de suas aplicações sem repensar arquitetura, processos ou modelos operacionais.

Com GenAI existe um risco parecido, pois simplesmente inserir um componente de AI sobre um processo existente não significa necessariamente que o processo tenha sido transformado, da mesma forma que automatizar uma atividade ineficiente pode apenas fazer com que aquela ineficiência seja executada mais rapidamente.

Talvez a pergunta mais importante não devesse ser “onde podemos colocar AI?”, mas sim “como este processo seria desenhado se estivesse sendo criado agora e tivesse desde o início acesso às atuais capabilities de AI?”.

Essa mudança é relevante porque permite questionar premissas existentes e evita que a tecnologia seja utilizada apenas para perpetuar processos desenvolvidos em um contexto completamente diferente.

Além disso, outra lição importante de Cloud está relacionada à necessidade de desenvolver capabilities organizacionais e não simplesmente adquirir tecnologia, pois FinOps, Cloud Architecture, Security, Platform Engineering e inúmeros outros conhecimentos precisaram amadurecer para que as organizações conseguissem operar adequadamente seus ambientes.

Com AI será provavelmente parecido, pois não bastará contratar modelos ou disponibilizar ferramentas, sendo necessário desenvolver AI Architecture, AI Engineering, AI Governance, AI Security, Model Evaluation, Data Engineering, Agent Management e outros componentes de uma nova disciplina operacional.

Realidade do impacto da GenAI comparada a outros hypes tecnológicos

Ao longo das últimas décadas, não faltaram tecnologias cercadas por enorme hype e previsões de transformação radical que posteriormente apresentaram resultados significativamente menores do que aqueles inicialmente projetados, sendo o metaverso um bom exemplo recente de como uma tendência extremamente comentada pode não

alcançar a velocidade de adoção que parte do mercado esperava.

GenAI parece apresentar características substancialmente diferentes porque seus benefícios podem ser experimentados diretamente e de forma relativamente simples por pessoas e organizações, sem a necessidade de esperar anos por mudanças de infraestrutura, comportamento ou ecossistema.

Uma pessoa pode utilizar AI para escrever código, analisar documentos, resumir conteúdos, estruturar apresentações, criar imagens, pesquisar informações ou auxiliar em inúmeras outras atividades praticamente imediatamente, algo que gera uma percepção de valor muito mais direta do que aquela encontrada em algumas ondas tecnológicas anteriores.

Essa experiência individual não significa automaticamente que os mesmos benefícios serão obtidos em escala empresarial, mas cria uma dinâmica extremamente poderosa de adoção bottom-up, na qual colaboradores passam a conhecer e utilizar as ferramentas antes mesmo que suas organizações tenham definido políticas, arquiteturas ou estratégias formais.

Isso ajuda a explicar por que Shadow AI se tornou uma preocupação relevante e por que simplesmente proibir o uso pode não representar uma estratégia sustentável, pois usuários que percebem ganhos reais de produtividade tendem naturalmente a buscar caminhos alternativos quando as ferramentas corporativas não atendem às suas necessidades.

A diferença fundamental, portanto, parece estar no fato de que GenAI já demonstrou utilidade concreta em uma variedade enorme de atividades, mesmo que parte do hype em torno de determinadas aplicações, previsões ou vendors inevitavelmente seja corrigida ao longo do tempo.

É perfeitamente possível assumir simultaneamente duas coisas que, à primeira vista, parecem contraditórias: existe hype excessivo em torno de diversos aspectos da AI e, ao mesmo tempo, a tecnologia possui potencial genuinamente transformador.

O desafio das empresas está justamente em separar uma coisa da outra.

Da experimentação para a industrialização

Talvez uma das mudanças mais importantes que deverá ocorrer nos próximos anos seja a transição de uma fase predominantemente experimental para uma etapa de industrialização da Inteligência Artificial dentro das empresas, algo que exigirá processos, plataformas e mecanismos de gestão muito mais maduros.

Durante uma etapa inicial, faz sentido permitir experimentação distribuída, testar

diferentes modelos, realizar Proofs of Concept e entender onde a tecnologia pode produzir valor, pois uma organização que tentar antecipadamente estabelecer um modelo excessivamente rígido corre o risco de impedir o próprio aprendizado necessário para construir sua estratégia.

Entretanto, chega um momento em que centenas de experimentos precisam ser consolidados em um portfolio coerente, permitindo identificar o que deve ser abandonado, aquilo que deve continuar como uso pontual e aquilo que efetivamente merece ser transformado em uma capability empresarial.

Essa passagem do experimento para produção deveria considerar não apenas a qualidade técnica da solução, mas também arquitetura, segurança, custos, dados, operação, suporte, integração, Observability, continuidade, compliance e ownership.

É justamente nessa etapa que muitas organizações podem encontrar dificuldades, pois aquilo que funciona muito bem em um ambiente controlado pode apresentar desafios completamente diferentes quando colocado em escala.

Uma solução utilizada por cinquenta pessoas não possui necessariamente as mesmas características econômicas ou operacionais quando passa para cinquenta mil, assim como um agente que executa dez atividades por dia não apresenta o mesmo perfil de risco quando passa a executar dezenas de milhares.

Da mesma maneira, um modelo que acessa uma pequena base de documentos não possui os mesmos desafios de segurança quando passa a consultar grande parte do conhecimento corporativo, fazendo com que industrializar AI signifique, em essência, transformar experimentos promissores em serviços administráveis e sustentáveis.

A ascensão da AI Governance

Outro ponto que se tornou muito mais relevante é o crescimento da AI Governance como uma disciplina própria, deixando progressivamente de ser um conjunto de princípios conceituais para se transformar em mecanismos técnicos e operacionais capazes de controlar o uso da tecnologia.

Inicialmente, muitas organizações começaram criando políticas relativamente simples para definir quais ferramentas poderiam ser utilizadas, quais tipos de informação poderiam ser compartilhados e quais situações exigiriam maior nível de aprovação, algo absolutamente necessário enquanto primeira camada de proteção.

Com o avanço da tecnologia, entretanto, Governance precisa ser operacionalizada, sendo necessário conhecer quais modelos estão sendo utilizados, quais agentes existem, quem é responsável por eles, quais dados acessam, quais ações podem

executar, qual é o respectivo nível de risco e como o comportamento dessas soluções pode ser monitorado ao longo do tempo.

Esse movimento tende a ganhar ainda mais importância conforme a quantidade de agentes cresce e as empresas percebem que não será viável administrar milhares de componentes autônomos através de processos manuais, documentos ou planilhas isoladas.

Nesse contexto, Model Inventory, Agent Inventory, Identity Management, Policy Enforcement, Continuous Monitoring, Auditability, Risk Classification e Lifecycle Management começam a formar uma nova camada de controles que provavelmente será incorporada ao modelo normal de gestão de tecnologia.

Governance não deveria ser percebida como uma forma de impedir inovação, mas exatamente como aquilo que torna possível ampliar a utilização da tecnologia sem perder o nível necessário de confiança.

Enterprise Architecture ganha ainda mais importância

Quanto mais AI se torna parte do ambiente corporativo, mais evidente fica que o tema não deveria permanecer isolado dentro de áreas especializadas de Data Science ou Innovation, pois a tecnologia começa a afetar simultaneamente Application Architecture, Data Architecture, Integration Architecture, Security Architecture e Infrastructure Architecture.

Quais modelos podem ser utilizados?

Existirá um AI Gateway corporativo?

Como será realizada a seleção entre modelos?

Quais aplicações poderão utilizar modelos externos?

Como serão tratados modelos open-weight?

Como agentes acessarão APIs e ferramentas?

Como identidade será controlada?

Como contexto empresarial será fornecido?

Como evitar vendor lock-in?

Essas perguntas pertencem claramente à agenda de Enterprise Architecture.

Isso é particularmente importante porque o mercado de AI ainda está evoluindo

rapidamente e uma decisão arquitetônica excessivamente acoplada a um fornecedor pode se tornar problemática caso novos modelos, padrões ou plataformas passem a oferecer vantagens significativas pouco tempo depois.

Flexibilidade arquitetural não significa evitar decisões, mas sim tomar decisões suficientemente conscientes para preservar alternativas futuras sempre que isso fizer sentido para a organização.

Ao mesmo tempo, não se pode cair no extremo oposto de uma arquitetura excessivamente abstrata e genérica construída apenas para garantir portabilidade teórica, mas que introduza complexidade adicional sem benefício real, fazendo com que o equilíbrio entre standardization e optionality provavelmente seja um dos grandes desafios arquiteturais desse novo cenário.

Data e contexto empresarial

Talvez um dos aspectos mais interessantes da evolução da AI seja que, apesar de toda a sofisticação dos modelos, algumas das questões mais antigas relacionadas a Data continuam absolutamente presentes e, em muitos casos, tornam-se ainda mais importantes.

Um Foundation Model pode possuir conhecimento sobre uma quantidade extraordinária de assuntos e ainda assim não conhecer adequadamente os produtos específicos de uma empresa, seus clientes, contratos, procedimentos, políticas, sistemas, estruturas organizacionais, exceções operacionais e todas as particularidades que formam seu contexto empresarial.

É justamente nesse ponto que o valor de tecnologias como Retrieval Augmented Generation, Knowledge Bases, Semantic Layers, Vector Search e mecanismos de Enterprise Context começa a se tornar evidente, pois não basta ter um modelo inteligente se ele não consegue compreender adequadamente a realidade específica da organização.

Data Quality continua importando.

Data Governance continua importando.

Metadata continua importando.

Data Lineage continua importando.

Knowledge Management continua importando.

Na realidade, AI pode aumentar a importância de todas essas disciplinas porque amplia significativamente a quantidade de formas pelas quais dados empresariais

podem ser utilizados para gerar decisões, recomendações e ações.

O antigo conceito de “garbage in, garbage out” continua absolutamente válido, embora agora exista a possibilidade de que o garbage seja apresentado através de uma resposta extremamente bem estruturada e convincente, algo que torna ainda mais importante assegurar qualidade e confiança sobre o contexto utilizado.

O equilíbrio entre autonomia e controle

Agentic AI adiciona ainda outra dimensão ao debate, pois quanto maior a capacidade concedida a um agente, maior precisa ser a clareza sobre aquilo que ele está autorizado a fazer e quais mecanismos serão acionados caso ultrapasse os limites esperados.

Essa preocupação dificilmente poderá ser resolvida através de uma simples divisão entre agentes “permitidos” e “proibidos”, pois existem níveis intermediários de autonomia que exigem controles proporcionais.

Um agente responsável por buscar informações em documentos internos pode trabalhar com controles relativamente simples de autenticação, autorização e logging, enquanto um agente capaz de modificar dados financeiros deveria exigir um nível muito mais robusto de validação, monitoramento, aprovação e rollback.

Da mesma forma, pode existir uma progressão natural entre agentes que apenas observam determinadas condições, aqueles que recomendam decisões, aqueles que podem executar ações após uma aprovação humana e aqueles suficientemente maduros e controlados para atuar autonomamente dentro de determinados guardrails.

Essa abordagem permite equilibrar produtividade e risco sem aplicar o mesmo nível de controle a situações completamente diferentes e, principalmente, cria uma trajetória de maturidade na qual a autonomia pode ser progressivamente ampliada conforme confiança, previsibilidade e mecanismos de proteção sejam demonstrados.

Talvez um princípio básico para o mundo agentic seja justamente este: autonomia deve ser concedida de forma progressiva e proporcional ao nível de confiança efetivamente demonstrado pela solução.

A identidade dos AI Agents

Existe ainda outro tema que provavelmente ganhará muito mais relevância conforme Agentic AI se torna comum dentro das organizações: como tratar a identidade digital desses agentes?

Durante décadas, Identity & Access Management foi predominantemente estruturado ao redor de dois tipos de entidades, pessoas e aplicações, cada uma delas recebendo credenciais e permissões que deveriam estar alinhadas às suas respectivas necessidades.

AI Agents introduzem uma nova categoria que combina características dos dois mundos, pois são componentes de software que podem agir dinamicamente, interpretar situações e eventualmente utilizar diversas ferramentas para atingir determinado objetivo.

Isso significa que não será suficiente apenas fornecer ao agente uma credencial genérica com acesso amplo aos sistemas necessários, pois essa abordagem destruiria grande parte dos princípios de Least Privilege e Segregation of Duties que foram construídos ao longo de décadas.

Quais permissões determinado agente precisa?

Essas permissões deveriam ser permanentes ou concedidas apenas durante determinada execução?

Um agente pode delegar atividades para outro agente?

Até onde essa cadeia de autorização pode chegar?

Como identificar posteriormente qual agente realizou determinada ação?

Quem é o owner daquela identidade?

Como um agente é desativado quando deixa de ser necessário?

Essas perguntas começarão inevitavelmente a fazer parte da agenda de Cybersecurity e Identity Governance e demonstram mais uma vez como Agentic AI ultrapassa rapidamente a fronteira de um simples recurso de produtividade.

Cybersecurity em um novo contexto

Artificial Intelligence também amplia significativamente a superfície de preocupação relacionada à Cybersecurity, pois além dos riscos tradicionais associados a aplicações, infraestrutura, identidades e dados, surgem novas classes de ataques e vulnerabilidades especificamente relacionadas ao comportamento dos modelos.

Prompt Injection, Model Manipulation, Data Poisoning, Sensitive Information Disclosure, Excessive Agency, Insecure Tool Access e diferentes formas de exploração dos mecanismos utilizados para conectar modelos a ferramentas passam a fazer parte do cenário de risco.

A complexidade aumenta particularmente quando agentes possuem capacidade de acessar dados e posteriormente tomar ações com base naquilo que encontraram, criando a possibilidade de que conteúdos maliciosamente preparados influenciem o comportamento da própria automação.

Nesse cenário, Security by Design precisa ser aplicado desde a concepção dos casos de uso, definindo claramente quais fontes de dados podem ser utilizadas, quais ferramentas podem ser acessadas, quais atividades podem ser executadas e quais controles adicionais precisam existir para cenários de maior criticidade.

Também se torna fundamental garantir rastreabilidade suficiente para reconstruir aquilo que ocorreu durante uma execução, algo que pode exigir logs muito mais ricos do que aqueles utilizados tradicionalmente para aplicações determinísticas.

Se determinado AI Agent tomou uma decisão incorreta, talvez não seja suficiente simplesmente saber que uma API foi chamada, pois será necessário compreender qual contexto foi utilizado, qual modelo estava ativo, quais instruções foram fornecidas, quais ferramentas foram acionadas e qual sequência de eventos resultou naquela decisão.

Pessoas e a transformação do trabalho

Uma grande parte do debate público relacionado à AI continua concentrada sobre quais profissões desaparecerão ou quantas pessoas serão substituídas, mas acredito que a transformação possa ocorrer de maneira muito mais gradual e distribuída do que essa visão binária normalmente sugere.

Muitas funções provavelmente continuarão existindo, mas terão parcelas significativas de suas atividades alteradas, algumas sendo automatizadas, outras aceleradas e outras completamente delegadas para modelos ou agentes.

Talvez a unidade correta de análise não seja apenas o job, mas as tasks que compõem cada função.

Um profissional pode continuar sendo responsável por determinada atividade empresarial enquanto partes substanciais da coleta de informações, análise preliminar, documentação, geração de materiais ou execução operacional são realizadas com apoio de AI.

Isso pode modificar profundamente o perfil de skills esperado, fazendo com que pensamento crítico, capacidade de validação, conhecimento de negócio e habilidade para trabalhar em colaboração com sistemas inteligentes ganhem relevância adicional.

Também acredito que existe um risco pouco discutido relacionado à perda progressiva

de determinadas competências quando pessoas passam a delegar excessivamente atividades para a tecnologia, criando uma dependência que pode reduzir sua própria capacidade de avaliar se determinada resposta está correta.

AI Literacy, portanto, deveria representar algo muito mais amplo do que simplesmente saber utilizar prompts, pois é necessário compreender as capacidades da tecnologia, suas limitações, riscos, formas de validação e principalmente a responsabilidade que continua existindo quando uma pessoa decide utilizar determinado resultado.

AI precisa gerar valor mensurável

Por fim, toda essa discussão precisa voltar para uma pergunta relativamente simples: qual valor está sendo criado?

Durante uma primeira fase tecnológica é natural que experimentação e aprendizado sejam considerados benefícios por si só, mas uma organização não deveria permanecer indefinidamente financiando iniciativas apenas porque elas utilizam uma tecnologia considerada estratégica.

Chega um momento em que será necessário demonstrar aumento de receita, redução de custos, ganho de produtividade, diminuição de riscos, melhoria de Customer Experience, redução de Cycle Time ou qualquer outro outcome suficientemente claro para justificar a continuidade do investimento.

Número de usuários utilizando uma ferramenta não representa necessariamente valor.

Quantidade de agentes criados também não.

Volume de tokens consumidos muito menos.

É necessário medir aquilo que mudou na operação ou no resultado empresarial, algo que exige sair progressivamente das métricas de adoção tecnológica e avançar para indicadores capazes de demonstrar impactos concretos.

Productivity Gain, Revenue Uplift, Cost Reduction, Cycle Time, Automation Rate, Human Intervention Rate, Accuracy, Reliability, Agent Success Rate, Cost per Successful Execution e Business Value Delivered são apenas alguns exemplos de indicadores que podem ser utilizados dependendo da natureza da solução.

Essa talvez seja uma das mudanças mais importantes na maturidade da discussão sobre AI.

Sair da pergunta “quanto estamos utilizando?” para chegar em “quanto valor estamos produzindo?”.

Concluindo

A Generative AI representa uma das transformações tecnológicas mais relevantes dos últimos anos e tudo indica que ainda estamos observando apenas as primeiras consequências de uma evolução que tende a avançar rapidamente em capacidade, autonomia, integração e presença dentro do ambiente empresarial.

O que começou para muitas pessoas como uma experiência surpreendente de conversar com um chatbot evoluiu rapidamente para geração de conteúdo multimodal, desenvolvimento de software, pesquisa, análise de documentos e uma quantidade crescente de aplicações que começam a transformar a forma como determinadas atividades são executadas.

Agora, com Agentic AI, a fronteira se move novamente e sistemas deixam progressivamente de apenas produzir respostas para começar a executar atividades, introduzindo uma nova geração de desafios relacionados à governança, identidade, segurança, arquitetura, accountability e controle operacional.

Isso reforça uma percepção que considero especialmente importante: a parte mais difícil da Inteligência Artificial talvez não seja simplesmente obter acesso aos melhores modelos.

Os melhores modelos poderão estar disponíveis para milhares de empresas ao mesmo tempo.

O verdadeiro diferencial estará na capacidade de cada organização construir ao redor deles uma estrutura capaz de transformar potencial tecnológico em resultados concretos, integrando arquitetura, dados, governança, segurança, processos, pessoas, Operating Model, disciplina financeira e gestão de riscos dentro de uma abordagem coerente.

É justamente nesse ponto que as discussões relacionadas à Artificial Intelligence começam a se parecer menos com um tema isolado de tecnologia e mais com qualquer grande transformação empresarial.

Não acredito que o caminho correto seja desacelerar por medo, especialmente porque a velocidade de evolução provavelmente tornaria uma postura excessivamente conservadora cada vez mais difícil de sustentar, assim como não acredito que simplesmente correr atrás de toda novidade possa ser considerado estratégia.

O equilíbrio está em experimentar rapidamente, aprender continuamente e industrializar seletivamente aquilo que efetivamente demonstra capacidade de gerar valor.

Creio que essa seja a questão central para CIOs e demais executivos ao longo dos próximos anos.

Não simplesmente perguntar se a organização utiliza Artificial Intelligence, mas perguntar se ela sabe por que está utilizando, onde essa tecnologia realmente produz resultados, quanto custa para operar, quais riscos está criando, como será governada e de que maneira se conecta à estratégia empresarial.

Se essas perguntas forem respondidas adequadamente, a Inteligência Artificial poderá se transformar em uma das mais poderosas alavancas de produtividade, inovação e transformação já disponibilizadas às empresas.

Caso contrário, existe sempre o risco de construir uma enorme coleção de soluções extremamente sofisticadas, tecnologicamente impressionantes e economicamente difíceis de justificar.

E, no fim das contas, essa continua sendo uma das distinções mais importantes do mundo de TI.

A diferença entre adotar tecnologia e efetivamente transformar tecnologia em valor.