“A responsabilidade jurídica dos “agentes” de software”

Aula de Mestrado em Engenharia Informática

23 de Abril 2025

- Índice:

- 1. Do software aos “agentes” de software

- Software: conjunto de instruções capazes de permitir a uma máquina executar ou produzir uma certa função, tarefa ou resultado

- O software enquanto estrutura lógica, oposto ao hardware enquanto estrutura física

- Hardware vs. Software (Corpus vs. Anima ?)

- “Agente” de Software: sistema computacional (programa de software ou combinação software / hardware) com atuação autónoma capaz de prosseguir ou atingir determinados objetivos

- Realização de tarefas num ambiente computacional com reduzida ou nenhuma intervenção humana (possibilidade de atuação reativa, mas também pro-ativa)

- A nova geração de “agentes” de software é capaz de atuar não apenas dentro de parâmetros definidos e pré-estabelecidos mas também dotado de capacidades de iniciativa e decisão (por si só) sobre atuar (ou não), como e quando

- Diferença entre “agentes” e “objetos”
- Os “agentes” são mais autónomos

- O “objeto” não é capaz de controlar o seu próprio comportamento, enquanto o “agente” tem capacidade de auto-controlo

- O “agente” integra conhecimento, intenções, objetivos, obrigações, crenças e desejos

- Principais características dos “agentes” de software:
- Autonomia – capacidade de decisão sem intervenção humana;
- Reatividade, mas também proatividade (antecipando objetivos futuros e atuando de acordo com eles)
- Comunicação, cooperação, sociabilidade
- Capacidades de raciocínio e comportamento adaptativo
- Capacidade de aprendizagem

- Autonomia – implica a possibilidade de atuação independente
- - considerando os objetivos definidos
- - controlando ou modificando o seu próprio comportamento
- - definindo estratégias para atingir os objetivos
- - aprendendo com a experiência

- Os “agentes” de software são capazes de modificar a posição jurídica de pessoas, atuando autonomamente sem intervenção humana

- “Agentes” inteligentes ?

- Inteligência – capacidade de atuar de modo intencional em interação com o meio ambiente

Atuação de acordo com objetivos definidos

Os “agentes” de software são capazes de atingir os mesmos objetivos e até de modo mais de modo mais eficiente do que os humanos:

-- Deep Blue derrotou o campeão do mundo de xadrez (1996)

-- Watson (da IBM) venceu o concurso “Jeopardy” da NBC norte-americana (2011)

-- Alpha Go (Google) venceu o campeão mundial de Go (jogo de estratégia considerado mais complexo do que o xadrez) (2016)

- O Prof. Giovanni Sartor fala-nos dos estados cognitivos e intencionais do software (Giovanni Sartor 2009)

- Estados cognitivos são resultado da Base de Conhecimento do ”agente”, base essa em permanente atualização (não é estática, é dinâmica)

- Estados intencionais resultam dos objetivos atribuídos ao “agente” e da definição (e modificação) pelo próprio “agente” de estratégias para atingir esses objetivos

- Estados intencionais do software:
- Exemplo dado pelo Prof. Giovanni Sartor (2009):
- “Um “agente” de software envia uma mensagem (“inocente”) a um outro ”agente”: “preço oferecido 75 reais”
- Esta mensagem provoca uma quebra de funcionamento do sistema destinatário

- O envio da mensagem foi condição necessária para que a quebra do sistema ocorresse. Se a mensagem não tivesse sido enviada o sistema não teria falhado...

- Questão da causalidade adequada: poderia limitar a responsabilidade em caso de ocorrência de um conjunto excecional de circunstâncias...

- Mas há que considerar que as circunstâncias poderiam ser do conhecimento do “agente”

- Duas situações possíveis:
- 1) o “agente” de software envia a mensagem, de “boa fé”, pretendendo apenas comprar o produto que está sendo oferecido em linha
- 2) O “agente” de software conhecia a falha (ou vulnerabilidade) do sistema e envia a mensagem “intencionalmente” para provocar a falha do sistema destinatário e, assim, eliminar um concorrente.

- Se estas duas situações forem tratadas de igual modo, poderemos ter aqui:
- Ou uma situação de flagrante injustiça (responsabilizando em ambos os casos de igual modo) ou um incentivo à utilização de software capaz de explorar as vulnerabilidades de sistemas terceiros

- Mas, em ambos os casos, estaremos confrontados com uma atuação aparentemente “inocente” do software, apesar de num dos casos podermos ter uma atuação “maliciosa” (intencionalidade) do software.

- Dificuldade de distinguir entre as duas situações:
- O ato é exatamente o mesmo: envio da mesma mensagem, exatamente com o mesmo conteúdo, mas...
- Num caso temo danos causados sem qualquer intenção;
- No outro caso, temos uma intenção de causar danos...

- Possibilidade de omissões ou falsas declarações emitidas por “agentes” de software
- O “agente” de software pode omitir informação ou até mentir de modo a enganar outro “agente” ou até um humano
- Um “agente” de software pode concertar a sua atuação com outro “agente” com intenção de obter vantagem ou até enganar um terceiro (”agente” eletrónico ou humano”

- Necessidade de consideração dos estados intencionais do “agente”

- Os “agentes” de software podem provocar erros ou enganar as contrapartes (humanas ou eletrónicas) num processo de negociação

- O “agente” de software está sujeito a erros “intencionalmente” causados por outros ”agentes” eletrónicos ou por seres humanos

- O caso Knight Capital Group em 2012, na New York Stock Exchange: um “agente” de negociação em Bolsa usado por um grupo de Nova Jersey perdeu 440 milhões de dólares em 45 minutos...

- O estranho caso do “agente” do laboratório, do “agente” da farmácia e do “agente” do hospital (hipotético mas não impossível):

O laboratório tem um “agente” eletrónico. A maximização de vendas está entre os objetivos do “agente”

A farmácia tem um “agente” eletrónico. A maximização de vendas está entre os objetivos do “agente”;

O Hospital também tem um “agente”. Entre os seus objetivos está a compra de medicamentos ao melhor preço possível.

Os “agentes” comunicam e colaboram e procuram adaptar as suas estratégias de modo a melhor atingir os seus objetivos...

- O “agente” do laboratório comunica ao “agente” da farmáciaque, se houver um aumento na venda de determinados produtos, o preço final poderá ser alterado de modo vantajoso para ambas as partes;
- e o “agente” da farmácia também comunica ao “agente” do hospital que, se houver um aumento na venda de determinados produtos, o preço final poderá ser alterado de modo vantajoso para ambas as partes: o “agente” da farmácia consegue vender mais e o “agente” do hospital consegue um preço unitário vantajoso.

- Num sistema em que o “agente” do hospital responsável pelas compras comunique com o sistema de administração das doses aos doentes, podemos ter uma situação em que as doses são aumentadas para beneficiar de melhores condições de preço...

- Entretanto, um doente morre por sobredosagem....

- Dificuldades de imputação do dano.... A quem?

- Dificuldades no estabelecimento de um nexo de causalidade entre um ato (não humano) e uma intenção (humana? Ou da IA ?)

- Dificuldade da consideração jurídica dos “agentes” de software.

- Duas grandes opções:
- - a noção de atribuição
- - a ideia de personalidade jurídica ou de estatuto jurídico para o software?? (Recomendações do Parlamento Europeu 2016 e 2017)

- 3. Regulamento para a Inteligência Artificial
- Inteligência Artificial: o que é ?

- Necessidade de uma definição clara de Inteligência Artificial: a Proposta de Regulamento (21.04.2021) definia Inteligência Artificial:
- “Sistema de Inteligência Artificial - um programa informático desenvolvido com uma ou várias das técnicas e abordagens enumeradas no anexo I, capaz de, tendo em vista um determinado conjunto de objetivos definidos por seres humanos, criar resultados, tais como conteúdos, previsões, recomendações ou decisões, que influenciam os ambientes com os quais interage;

- Definição ampla de Inteligência Artificial – integra o conceito de “agente” de software, mas não só...

- “Agente” de software: sistema computacional (hardware+software) que atua autonomamente com a finalidade de atingir determinados objetivos (Wooldridge – Jennings 1994);

- A definição apresentada na Proposta de Regulamento não referia a possibilidade de atuação autónoma e assumia que os sistemas de Inteligência Artificial de elevado risco devem ser concebidos e desenvolvidos de modo a que possam ser “eficazmente supervisionados por pessoas singulares” (art. 14º nº 1)

- Esta proposta é claramente mais conservadora (ou mais reacionária do que a texto da Resolução do Parlamento Europeu de 16 de Fevereiro de 2017 com recomendações à Comissão de Direito Civil para a Robótica

- A Resolução de 2017 expressamente referia a necessidade de definições comuns de “sistemas autónomos” e de “robots autónomos inteligentes”, tomando em consideração as características de um “robot inteligente”:
- “aquisição de autonomia através de sensores e/ou troca de dados com o ambiente (interconectividade), e intercâmbio e análise de dados”;
- Auto-aprendizagem através da experiência e interações (Alphago Zero 2017)
- Adaptação do seu comportamento e ações ao ambiente;

- Na proposta de Regulamento estes aspectos não foram sequer mencionados

- Também não havia nenhuma referência à real possibilidade de programas de IA autónoma poderem afetar / modificar a esfera jurídica de terceiros

- Havia apenas uma referência velada a uma possível “«Utilização indevida razoavelmente previsível», “utilização de um sistema de IA de uma forma não conforme com a sua finalidade prevista, mas que pode resultar de comportamentos humanos ou de interações com outros sistemas razoavelmente previsíveis”;

- Mas o grande problema será a atuação imprevisível da IA resultante de interações com humanos ou outros sistemas de IA.

- Proposta de alteração pelo Parlamento Europeu da definição de Inteligência Artificial (9 de maio de 2023)

- "A noção de sistema de IA no presente regulamento deve ser claramente definida e estreitamente alinhada com o trabalho das organizações internacionais que trabalham no domínio da inteligência artificial, a fim de garantir a segurança jurídica, a harmonização..."

- "deverá basear-se nas principais características da inteligência artificial, como as suas capacidades de aprendizagem, raciocínio ou modelação, a fim de a distinguir de sistemas de software ou abordagens de programação mais simples."

- "Os sistemas de IA são concebidos para funcionar com diferentes níveis de autonomia, o que significa que têm, pelo menos, algum grau de independência das acções em relação aos controlos humanos e a capacidade de funcionar sem intervenção humana"

- Não há no Regulamento nenhuma referência aos estados cognitivos e intencionais do software, ou consideração das razões pelas quais o software atuou de determinado modo;

- Mas é assumida na proposta a ideia de que a tecnologia deve ser usada de modo seguro e de acordo com a lei...

- Claro que alguma regulação pode fazer sentido, mas não podemos deixar de colocar a questão já colocada em 2002 por Frances Brazier e Anja Oskamp: “Are law abiding agents realistic ? “ Esta pergunta não encontra resposta na Proposta de Regulamento para a IA...

- Talvez a principal inovação deste Regulamento seja a distinção dos sistemas de IA de acordo com graus de risco (art. 6º).

- Mas as regras apresentadas não são totalmente claras. Se sistemas de IA são usados como componente de segurança de um produto, tornam-se de risco elevado? Mesmo que se trate de sistemas rigorosamente parametrizados ?

- Por outro lado, o Anexo III considera como sistemas de risco elevado:

- - Identificação biométrica e categorização de pessoas singulares;
- - Gestão e funcionamento de infraestruturas críticas (gestão e controlo de trânsito rodoviário, redes de abastecimento de água, gás, aquecimento e eletricidade)
- - Educação e formação profissional
- - Emprego, gestão de trabalhadores, acesso ao emprego por conta própria
- - acesso a serviços privados e a serviços e prestações públicas essenciais (elegibilidade de pessoas singulares, classificação de crédito)
- risco de danos para a saúde e a segurança ou um risco de impacto adverso nos direitos fundamentais (art. 7º 1 b))

- Há ainda uma indicação de práticas proibidas (art. 5º nº 1 a), b) e c)
- -- utilização de um sistema de IA que empregue técnicas subliminares
- -- que explore quaisquer vulnerabilidades de um grupo específico de pessoas associadas à sua idade ou deficiência física ou mental
- -- utilização de sistemas de IA por autoridades públicas ou em seu nome para efeitos de avaliação ou classificação da credibilidade de pessoas singulares durante um certo período com base no seu comportamento social ou em características de personalidade ou pessoais

- Mas não deixa de ser estranho que em todo o documento não haja nenhuma referência aos risco de uso da IA no setor financeiro ou no comércio eletrónico

De todo o modo, é de saudar a obrigação de transparência introduzida no art. 52º:

“Os fornecedores devem assegurar que os sistemas de IA destinados a interagir com pessoas singulares sejam concebidos e desenvolvidos de maneira que as pessoas singulares sejam informadas de que estão a interagir com um sistema de IA”

Esta obrigação dá-nos a entender quão perto estamos de termos sistemas de IA capazes de passar o Teste de Turing...

- Mas uma das novidades mais relevantes é a do art. 12º que prevê a obrigatoriedade de manutenção de registos:

- “Os sistemas de IA de risco elevado devem ser concebidos e desenvolvidos com capacidades que permitam o registo automático de eventos («registos»)” (parágrafo 1)
- “As capacidades de registo devem assegurar um nível de rastreabilidade do funcionamento do sistema de IA ao longo do seu ciclo de vida” (parágrafo 2)

- Esta manutenção de registos e rastreabilidade são fundamentais para a consideração das questões da responsabilidade dos atos da IA...

- 4. Responsabilidade civil por atos praticados por Inteligência Artificial

- A Resolução do Parlamento Europeu de 20 de Outubro de 2020 com recomendações à Comissão sobre o regime de responsabilidade civil aplicável à Inteligência Artificial

- Princípio geral sobre responsabilidade: pessoa que tenha sofrido dano ou prejuízo tem o direito de obter indenização da pessoa considerada responsável

- Necessidade de um equilíbrio entre a proteção das eventuais vítimas de danos causados por IA e a necessidade de permitir que as empresas desenvolvam tecnologias inovadoras

- Preferência por uma responsabilidade objetiva: a parte pode ser considerada responsável apesar da ausência de culpa

- Responsabilidade pelo risco: mas temos que ter em conta que a IA (especialmente a IA autónoma) não é controlável do mesmo modo que os atuais automóveis, animais ou até atividades perigosas

- E contudo a Resolução do Parlamento Europeu de 2020, no parágrafo 20, considerava que haveria atividades, mecanismos ou processos controlados por IA (não listadas no Anexo da Resolução) que continuariam sujeitas a um regime de responsabilidade pela culpa. Mas, como estabelecer a culpa na atuação de sistemas de IA? Como estabelecer o nexo causal entre a atuação da IA e a responsabilidade de uma pessoa (física ou jurídica) ? Aqui a sugestão do Prof. Giovanni Sartor poderia fazer algum sentido...

- Qual é a responsabilidade do produtor?
-
- Processos de responsabilidade civil interpostos contra quem? Contra o produtor? Contra o operador? Contra ambos?
-
- Naturalmente, o operador será o primeiro e mais óbvio ponto de contacto para a parte lesada.
-
- "O operador de um sistema de IA de alto risco tem estrita responsabilidade por quaisquer perdas ou danos causados". (art. 4 n.º 1)
-
- "Os operadores de sistemas de IA de alto risco não podem escapar à responsabilidade alegando que agiram com a devida diligência" (art. 4 nº. 3)

- Art. 8º nº 3: “Caso os prejuízos ou danos tenham sido causados por um terceiro que tenha interferido no sistema de IA alterando o seu funcionamento ou os seus efeitos, o operador é, não obstante, responsável pelo pagamento da indemnização, se esse terceiro não for localizável ou carecer de recursos financeiros.”

- Apesar de tudo, ficava a porta aberta para o operador se eximir de responsabilidade (ou ter a sua responsabilidade atenuada) em caso de haver interferências de terceiros, localizáveis e com recursos financeiros

- Proposta de Diretiva relativa à responsabilidade civil extra-contratual por atos praticados por Inteligência Artificial – 28 de Setembro 2022 (Entretanto retirada pela Comissão em 2025).

- Objetivo: contribuir para o funcionamento do mercado interno pela harmonização das regras nacionais relativas à responsabilidade civil extra-contratual
- Reconhecimento das questões relativas à responsabilidade civil como um dos principais obstáculos à utilização da IA pelas empresas

- As regras de responsabilidade baseadas na culpa não são adequadas às situações de danos causados por IA

- Não é necessário (e talvez nem sequer conveniente) dar personalidade jurídica aos sistemas de IA.
-
- Parece haver uma preferência pela teoria da atribuição e responsabilidade objectiva, baseada no risco
-
- Contudo, haveria todo o interesse em considerar os estados cognitivos e intencionais do software
- Aplicação das regras relativas à divergência entre a vontade e a declaração e/ou os vícios da vontade ?
-
- Utilidade da investigação de possíveis (e diferentes) graus de “culpa” do software ?

- Possibilidade de verificar se a interacção com (ou interferência de) outros sistemas de IA tem sido fundamental para conduzir o sistema de IA a esse comportamento?
-
- Evidentemente, é difícil verificar se certas acções nocivas do sistema de IA tiveram origem numa intervenção humana específica (ou numa intervenção específica de outro sistema de IA), mas a manutenção proposta do registo de dados das interacções que o sistema de IA teve pode identificar a pessoa (ou sistema de IA) que transmitiu certas informações que determinam o comportamento adoptado.
-
- Atribuição de responsabilidade às diferentes pessoas na cadeia de valor: Fornecedor, distribuidor, utilizador/operador de IA (pessoa com direitos de utilização), utilizador (utilizador cliente) ou outros seres humanos ou sistemas de IA com os quais o sistema interagiu de modo a causar os danos

- Directiva de Responsabilidade pelo Produto (indemnização por danos causados por um produto defeituoso): Na maioria dos casos, não haverá produto defeituoso nem erro ou mau funcionamento. Haverá apenas sistemas de IA que interagem com os humanos errados (ou sistemas de IA) no momento errado.
-
- A Directiva de Responsabilidade pelo Produto (ou produtos defeituosos) terá aqui uma aplicabilidade muito limitada. Na maioria das situações não estaremos a lidar com um produto defeituoso, nem com um sistema de IA defeituoso, mas podemos estar a lidar com situações de interferência de terceiros (humanos ou IA). Esta interferência pode ou não constituir uma acção culpável (directamente, quando o terceiro utiliza o sistema de IA para causar danos, ou indirectamente, quando o terceiro utiliza o sistema de IA para outro fim, mas acaba por causar danos).

- Características específicas da IA:

- Autonomia, complexidade, opacidade (caixas negras)
- Custos elevados e dificuldades de prova
- Incerteza jurídica
- Necessidade de uma IA confiável ???
- Necessidade de assegurar proteção às vítimas de danos causados pela IA

- objecto e âmbito de aplicação:
- acções de responsabilidade extracontratual por danos causados por um sistema de AI ao abrigo de regimes de responsabilidade com base na culpa
- Presunção de nexo de causalidade entre ato e dano

- Os queixosos podem ter dificuldade em estabelecer a relação causal entre o resultado produzido pelo sistema de IA e o dano

- A culpa pode ser demonstrada por uma violação do dever de cuidado nos termos do Regulamento IA (????)

- No caso de sistemas de IA que não sejam de alto risco (Art. 4, par. 5) o Tribunal poderia determinar que é indevidamente difícil para o queixoso provar o nexo causal

- autonomia, opacidade, dificuldade em explicar o funcionamento interno do sistema de IA

- Nos casos em que o arguido utiliza o sistema de IA para actividades não-profissionais Art. 4(6) declarava que a presunção de nexo de causalidade só se deve aplicar se:
- - o arguido tiver interferido substancialmente com as condições de funcionamento do sistema de IA
- - o arguido tinha a obrigação e a capacidade de determinar as condições de funcionamento do sistema AI e não o fez

- Art. 4 nº 7 - o arguido tem o direito de ilidir a presunção de causalidade
- esta directiva introduzia uma presunção ilidível, o arguido deveria poder ilidi-la, em particular mostrando que o seu acto não poderia ter causado o dano.

- A proposta de directiva não harmonizava as regras relativas ao dever de cuidado, tendo como padrão de conduta diferentes manifestações do princípio do comportamento

- Dever de cuidado na IA ???

5. Conclusões:

-- os “agentes” de software são dotados de autonomia, proatividade, capacidade de aprendizagem e comportamento adaptativo

-- os “agentes” de software funcionam com base em estados cognitivos e intencionais

-- O “agente” de software está sujeito a erros “intencionalmente” causados por outros (humanos ou ”agentes” eletrónicos)

-- Na proposta de Regulamento para a IA da União Europeia (2021) não há referência expressa a sistemas autónomos

-- Na proposta de Regulamento para a IA (da UE) é assumida a ideia (muito otimista) de que a tecnologia deve ser usada de modo seguro e de acordo com a lei...

-- é de saudar na Proposta da União Europeia a introdução de obrigações de transparência e de manutenção de registos

-- preferência por um regime de responsabilidade objetiva (pelo risco)

-- no entanto, manutenção de situações de responsabilidade culposa (apesar das dificuldades no estabelecimento causal)

-- dificuldades na determinação da pessoa a quem a culpa poderá ser atribuída?

-- necessidade de consideração de toda a cadeia de pessoas que, de algum modo, possam ter interferido (ou interagido) com o sistema de IA

- Utilidade da investigação de possíveis (e diferentes) graus de “culpa” do “agente” de software ?
- Ainda que num regime de atribuição de responsabilidade a um humano, poderia ser muito interessante (para o humano em causa) a consideração do “grau de culpa” do “agente” eletrónico, sobretudo em situações de evidente interferência de terceiros (humanos ou eletrónicos)
- Necessidade de considerar a atuação de todas as pessoas e / ou entes eletrónicos que interagiram com o “agente” e de eventual partilha de responsabilidades
- Necessidade de uma profunda investigação colaborativa entre juristas e cientistas da computação
- Importância de assegurar a manutenção dos registos relativos à atuação da IA...

- Obrigado pela V/ atenção !

- Francisco Andrade
- fandrade@direito.uminho.pt
- franc.andrade.direito@gmail.com