

Universidade do Minho

Escola de Engenharia

Departamento de Informática

Segurança, Confiança e Relações Jurídicas

Paulo Novais, Cesar Analide, Filipe Gonçalves

Perfil SI :: Agentes Inteligentes

Agenda

- Confiança
- Nível Individual
- Estratégias de Evolução e Aprendizagem
- Modelos de Reputação
- Modelos de Confiança Socio-Cognitivos
- Nível do Sistema
- Protocolos de Interacção
- Mecanismos de Reputação
- Mecanismos de Segurança
- Personalidade Jurídica?
- Conclusões



- Association for the Advancement of Artificial Intelligence:



www.aaai.org

- “Authors have agents... professional athletes have agents... movie stars have agents... and you have agents too. Because an agent is someone with expertise who is entrusted to go out and act on your behalf, the computer programs that help you to maximize your computing experiences are called ‘agents’. The next time that you search for specific information on the internet, picture your own agent or group of agents at work, with each knowing just what you’re interested in and how important your time is.”

Necessitamos...

- De ter confiança que os agentes fazem o que dizem... eles fazem;
- Estar confiantes de que a nossa privacidade está protegida;
- Acreditar que os riscos de segurança envolvidos em confiar em agentes que realizam operações em nosso nome são mínimos.



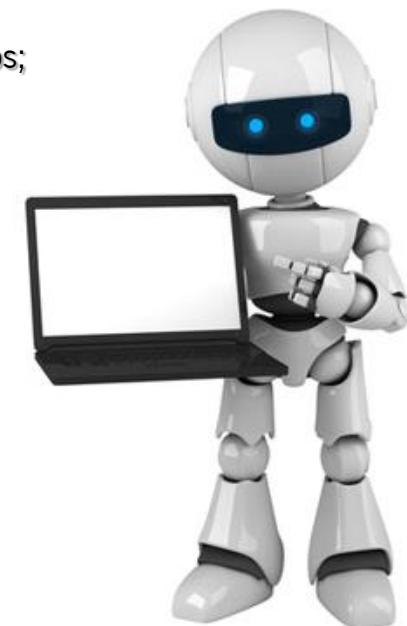
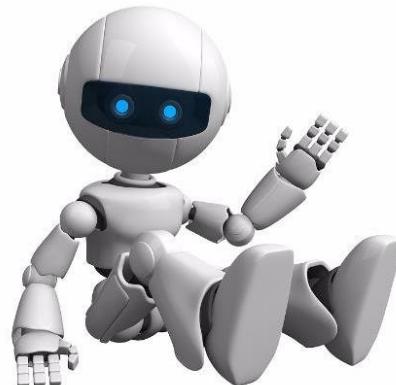
Confiança e Delegação de Poderes

- Se não há risco, a questão da confiança não se coloca;
- O ato de delegar pressupõe a confiança de passar a responsabilidade de uma tarefa para uma outra entidade (agente/humanos).



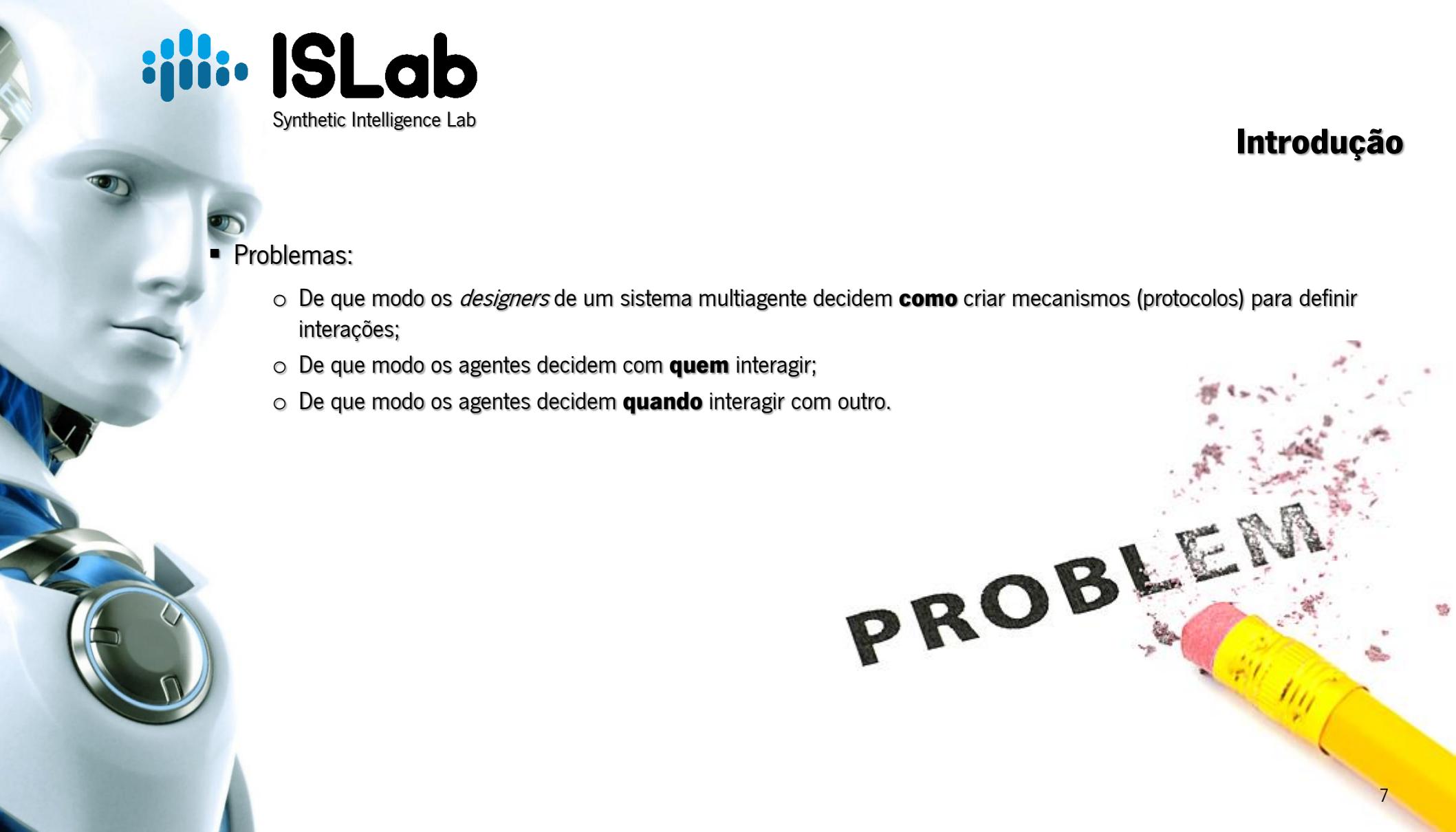
“This came in from somewhere. Could you
send it somewhere else?”

- Desafios na aplicação de agentes inteligentes (agentes de *software*) em sistemas distribuídos abertos/larga escala:
 - Agentes representam diferentes *stakeholders*;
 - Agentes podem entrar ou sair do sistema a qualquer momento;
 - Agentes de diferentes características podem entrar no sistema e interagir com outros;
 - Agentes podem trocar serviços e colaborar.



- Problemas:

- De que modo os *designers* de um sistema multiagente decidem **como** criar mecanismos (protocolos) para definir interações;
- De que modo os agentes decidem com **quem** interagir;
- De que modo os agentes decidem **quando** interagir com outro.

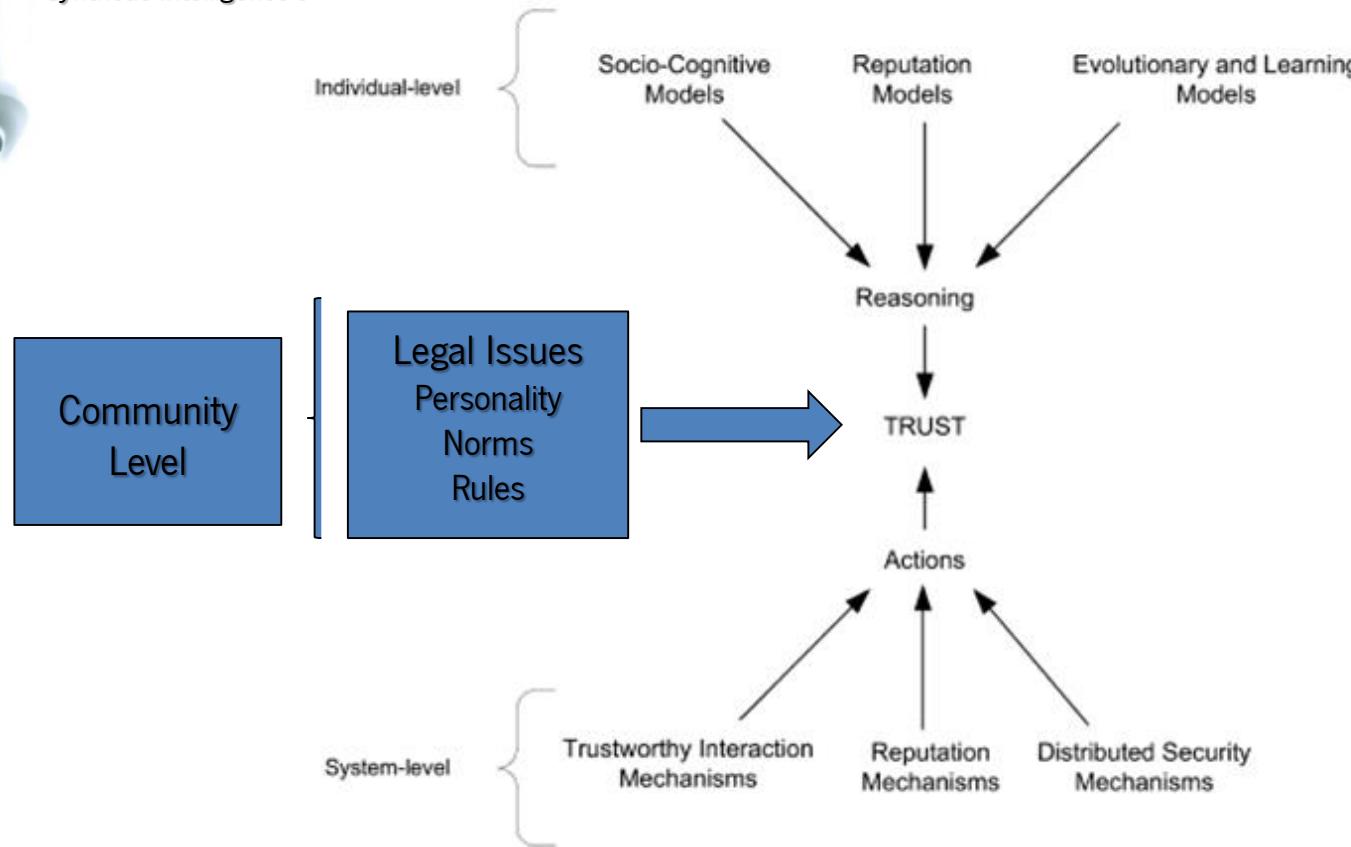


PROBLEMA

- “Trust is the belief that the other party will do what it says it will, reciprocate or given an opportunity, defect to get higher payoffs.”

“Dasgupta, P. 1998 Trust as a commodity. In Gambetta, D. (ed.), Trust: Making and Breaking Cooperative Relations. Blackwell, pp. 49–72.”





Adapted from

S. D. Ramchurn, D. Huynh and N. R. Jennings (2004) "Trust in multiagent systems" The Knowledge Engineering Review 19 (1)1-25.

Confiança ao Nível Individual

- 
- Os agentes podem:
 - Interagir com agentes e aprender o seu comportamento ao fim de algumas interações;
 - Interrogar outros agentes sobre a sua percepção dos potenciais parceiros;
 - Caracterizar as motivações conhecidas de outros agentes.
 - Através de múltiplas interações os agentes aprendem o que esperar de outros agentes, utilizando:
 - Modelos de confiança;
 - Aprendizagem:
 - Direta: métodos em que o agente possa aprender a detectar honestidade ou desonestidade;
 - Indireta: confiar noutrios na informação dada por outros agentes.

Modelos de Reputação

- Reputação: opinião de alguém sobre algo.
- Agentes avaliam o comportamento de outros agentes, em termos de:
 - Mal comportado → má reputação
 - Bem comportado → boa reputação
- Sistemas de reputação clássicos (p.ex., eBay e Amazon):
 - Recebem informações sobre grau de satisfação com as interações efetuadas.



Modelos de Confiança Socio-cognitivos

- 
- Os modelos anteriores são baseados na valorização de resultados de interações, mas também pode ser importante considerar a percepção subjetiva.
 - As crenças são essenciais para determinar a “quantidade” de confiança a colocar num agente:
 - Competência: uma avaliação positiva noutro agente onde é afirmado que esse agente é capaz de cumprir o que prometeu;
 - Esperança: acredita-se que o agente fará o que se propôs;
 - Persistência: acredita-se que o agente é suficientemente estável para concretizar o que se propôs fazer;
 - Motivação: acredita-se que o agente tem interesses pessoais para colaborar;
(muito utilizada para ganhos a longo termo)

Confiança (pessoal)

- Subjectiva e formada por um indivíduo baseado em crenças, observações, raciocínio, estereótipos sociais e experiências;
- Desenvolve-se na sequência de uma experiência positiva;
- Reduz-se em caso contrário;
- Existem diferentes disposições para a confiança.



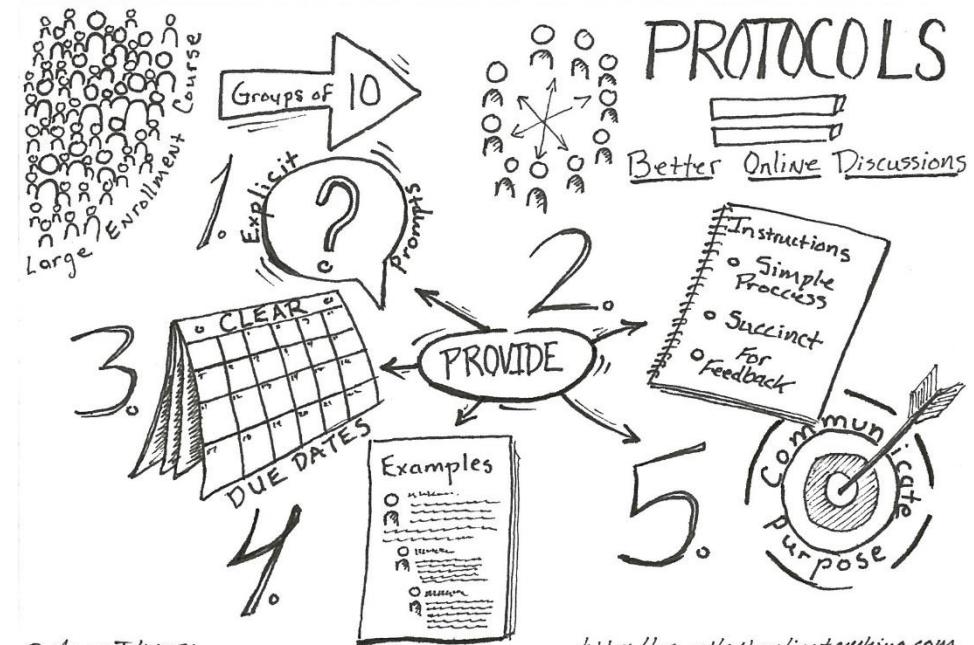
Confiança ao Nível do Sistema

- Os agentes presentes no sistema são obrigados a cumprir as suas regras;
- O sistema é um ser “todo poderoso”, que atribui e restringe regras;
- Deste modo, até os agentes egoístas têm de se submeter às regras estabelecidas;
 - Protocolos de interação;
 - Mecanismos de reputação;
 - Mecanismos de segurança que garantem que novos agentes no sistema possam ser confiáveis.



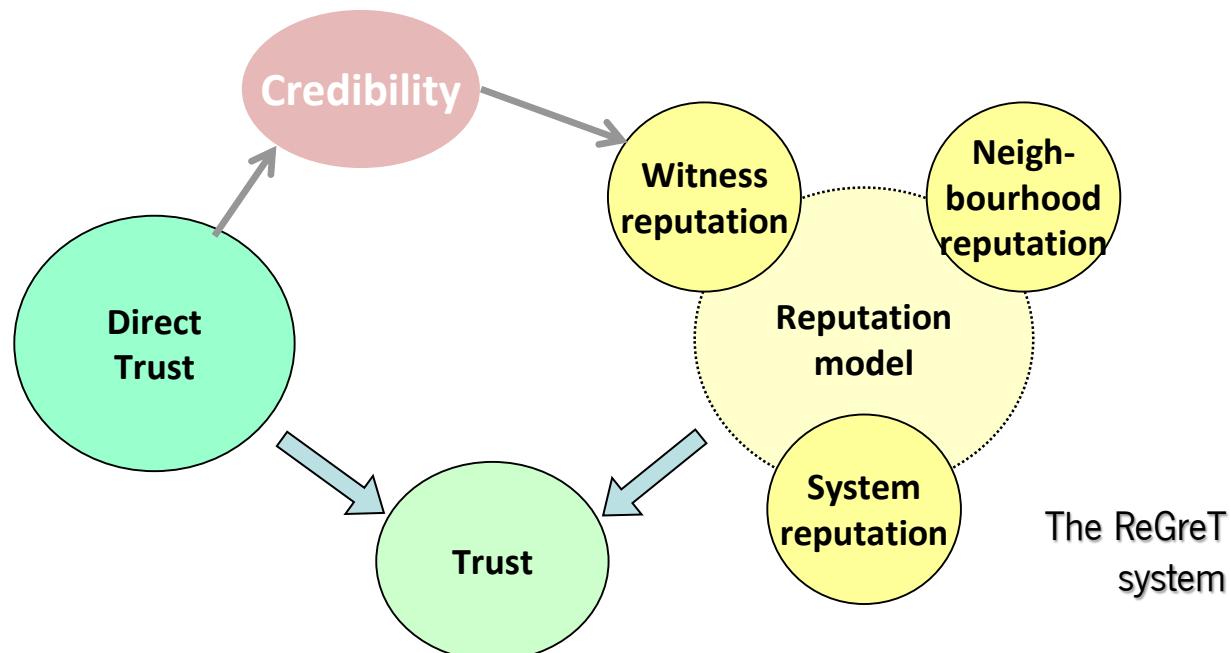
Protocolos de Interação

- Procuram prevenir que agentes mintam ou especulem enquanto interagem;
- Impôr regras que determinem os passos que devem ser seguidos;
- Regular a informação que pode ser revelada pelos agentes.



Mecanismos de Reputação

- Sabater J., Sierra C., Regret: Reputation and social network analysis in multi-agent systems.



Mecanismos de Segurança

- Todos os modelos de interação são baseados na premissa de que os agentes são reconhecidos pela sua identidade;
- Requisitos de segurança essenciais para agentes confiarem nas mensagens trocadas entre si:
 - Identidade;
 - Permissões de acesso;
 - Integridade do Conteúdo;
 - Privacidade do Conteúdo.



Confiança ao nível da Comunidade

- Considerar o agente como mero instrumento (como uma máquina)?
 - Considerar o processo declarativo como realizado por uma pessoa humana?
 - Ficção jurídica ou presunção jurídica?
 - Uma pessoa humana consente na emissão de uma declaração de vontade e na celebração de um contrato, apesar de nem ter conhecimento (ou nem sequer estar ciente) de que algo foi declarado?



Contratação eletrónica e agentes

- Poderemos aceitar aqui uma noção de atribuição?
“the operations of an intelligent agent are attributed to the human who uses the agent?”
(Weitzenboek, 2001)
- Será que o único consentimento válido e relevante é o da pessoa a favor (ou em nome) de quem o agente de *software* atua?
- Será que poderemos estabelecer uma relação entre a ação (não humana) e uma intenção humana?





- A vontade e o consentimento terão que ser inferidos (ou ficcionados) como vontade e consentimento do humano que utiliza o agente de *software* (ou em nome de quem o agente de *software* atua) ?
- Os riscos da atuação dos agentes de *software* recarriam, em termos de responsabilidade, no programador original, no proprietário ou no utilizador?
(as pessoas que programam, as que controlam e as que utilizam?)

Contratação eletrónica e agentes



THE PROBLEM ABOUT BEING A PROGRAMMER

My mom said:
"Honey, please go to the market and buy 1 bottle of milk. If they have eggs, bring 6"

I came back with 6 bottles of milk.

She said: "Why the hell did you buy 6 bottles of milk?"

I said: "BECAUSE THEY HAD EGGS!!!!"



- Uma pessoa que detém ou controla uma máquina é responsável pela utilização da máquina.
- Mas, será que existe aqui um controlo do agente? Será que programadores e utilizadores estarão em condições de antecipar todas as possibilidades de comportamento do agente de *software*?

Contratação eletrónica e agentes



THE PROBLEM ABOUT BEING A PROGRAMMER

My mom said:
"Honey, please go to the market and buy 1 bottle of milk. If they have eggs, bring 6"

I came back with 6 bottles of milk.

She said: "Why the hell did you buy 6 bottles of milk?"

I said: "BECAUSE THEY HAD EGGS!!!!"

Personalidade Jurídica

- Agentes de *software* como pessoas jurídicas;
- A atribuição de personalidade jurídica a agentes de *software* teria algumas vantagens evidentes:
 - Tornaria, desde logo, evidentemente válidas as declarações e contratos resultantes da atuação dos agentes de software;
 - Tornaria as questões de responsabilidade menos assustadoras para programadores, proprietários e utilizadores de agentes de *software*, pois se fosse considerada uma responsabilidade do agente de *software*, limitaria ou excluiria a responsabilidade dos humanos pelo comportamento do agente de *software*;



Personalidade Jurídica

- Mas existem evidentes dificuldades:

- Identificação dos agentes: o que constitui um agente de *software*? Apenas o *software*? O *hardware*?
- Mobilidade e ubiquidade dos agentes;
- Capacidade de clonagem dos agentes;
- Será possível estabelecer um domicílio para um agente?
- Obrigações patrimoniais: fará sentido atribuir um património a um programa de *software*?
- Pode um programa de *software* ser responsabilizado por atos ou omissões intencionais ?
- Pode-se falar de dolo e negligência de um agente de *software*?
- Pode um agente de *software* atuar de boa-fé ou de má-fé?
- Pode-se ação judicialmente um agente de *software*?



Personalidade Jurídica

- Uma pessoa jurídica não humana teria que ser constituída e registada;
- No caso da personalização jurídica de agentes de *software*:
 - Poderia ser-lhes atribuído um domicílio?
 - Deveria ser estabelecido um montante mínimo de capital obrigatório?
 - Deveria ser estabelecido um regime de seguro de risco obrigatório?
 - Como resolver juridicamente os litígios resultantes da atuação do agente de *software*?
 - Poderia ser representado em Tribunal?
 - Poderia ser alvo de uma ação executiva?



- 
- Poderemos recorrer ao mecanismo da representação?
 - Mas um representante tem de ter personalidade jurídica;
 - Uma criança até pode ser representante (ainda que limitada na sua capacidade jurídica) mas um programa de computador não pode;
 - E, no entanto, a capacidade de “raciocínio” e de representação das situações e consequências decorrentes de uma atuação até pode ser bem maior nos programas de *software* do que nas crianças!
 - Representação sem poderes e ratificação?
 - Um agente de *software* não pode ter poderes de representação, não porque não possa ter capacidade de querer e entender, mas porque não é uma pessoa jurídica:
 - Os agentes de *software* poderiam ser considerados representantes sem poderes?
 - O principal poderia *a posteriori* ratificar os negócios concluídos pelo “representante” eletrónico?

Contratação eletrónica e agentes de *software*

- Mas como trataria O Direito, em caso de litígio, uma declaração emitida por um programa de *software*?
- Inexistente?
- Pode uma declaração inexistente ser ratificada?

- E como encarar a subdelegação de tarefas entre agentes de *software*?
 - Os agentes de *software* são capazes de cooperar entre si, bem como de distribuir ou atribuir tarefas...

- Poderemos falar de subcontratação de agentes de *software*?
 - Mas os agentes de *software* não são pessoas jurídicas...

Contratação eletrónica e agentes de *software*

- 
- De todo o modo, a possibilidade de utilização de agentes de *software* (mais ou menos inteligentes) no comércio eletrónico é algo que já existe;
 - Deveremos aceitar a ficção de que eles são apenas instrumentos usados por humanos?
 - Deveremos considerá-los como pessoas jurídicas?
 - Poderemos encarar outras possibilidades?
 - Uma nova visão do contrato, já não considerado como um acordo de vontades mas como o resultado de atos de máquinas e mecanismos mais ou menos inteligentes?
 - Uma personalidade jurídica limitada?
 - **São questões em aberto...**



Conclusões

- 
- O sistema pode ser desenhado para forçar os agentes a serem confiáveis;
 - Utilizando os seus modelos de confiança, os agentes podem:
 - Raciocinar sobre estratégias usar;
 - Raciocinar sobre a informação recolhida;
 - Raciocinar acerca das motivações e capacidades dos parceiros de interação.
 - Os mecanismos e protocolos apresentados procuram forçar os agentes a agir e a interagir com confiança:
 - Impondo condições;
 - Utilizando a reputação para promover futuras interações;
 - Impondo padrões especificados de boa conduta.
 - As agentes com personalidade jurídica?

Referências

- Sabater J., Sierra C., Regret: Reputation and social network analysis in multi-agent systems. In Proceedings of First International Conference on Autonomous Agents and Multiagent Systems (AAMAS), pages 475–482, 2002.
- Dasgupta P. , Trust as a commodity. In Gambetta, D. (ed.), Trust: Making and Breaking Cooperative Relations. Blackwell, pp. 49–72, 1998.
- Ramchurn S., Huynh D., Jennings N., Trust in Multiagent Systems, The Knowledge Engineering Review 19 (1)1-25, 2004
- Sabater J., Sierra C., Review on computational trust and reputation models Artificial Intelligence Review ,24 (1) :33-60, 2005.

Referências

- Castelfranchi C., Falcone R., Principles of trust for MAS: Cognitive anatomy, social importance and quantification. Proc of ICMAS'98, pages 72-79, 1998.
- Castelfranchi C., Rosis F., Falcone R.: Social Attitudes and Personalities. In Agents, Socially Intelligent Agents, AAAI Fall Symposium Series 1997, MIT in Cambridge, Massachusetts, November 8-10, 1997.
- Andrade F., Novais P., Machado J., Neves J. Contracting Agents: legal personality and representation, Artificial Intelligence and Law, Volume 15, N° 4, ISSN 0924-8463, pp 357-373, Springer-Verlag, 2007.

Universidade do Minho

Escola de Engenharia

Departamento de Informática

Paulo Novais, Cesar Analide, Filipe Gonçalves

Perfil SI :: Agentes Inteligentes