



Rijksinspectie Digitale Infrastructuur
Ministerie van Economische Zaken

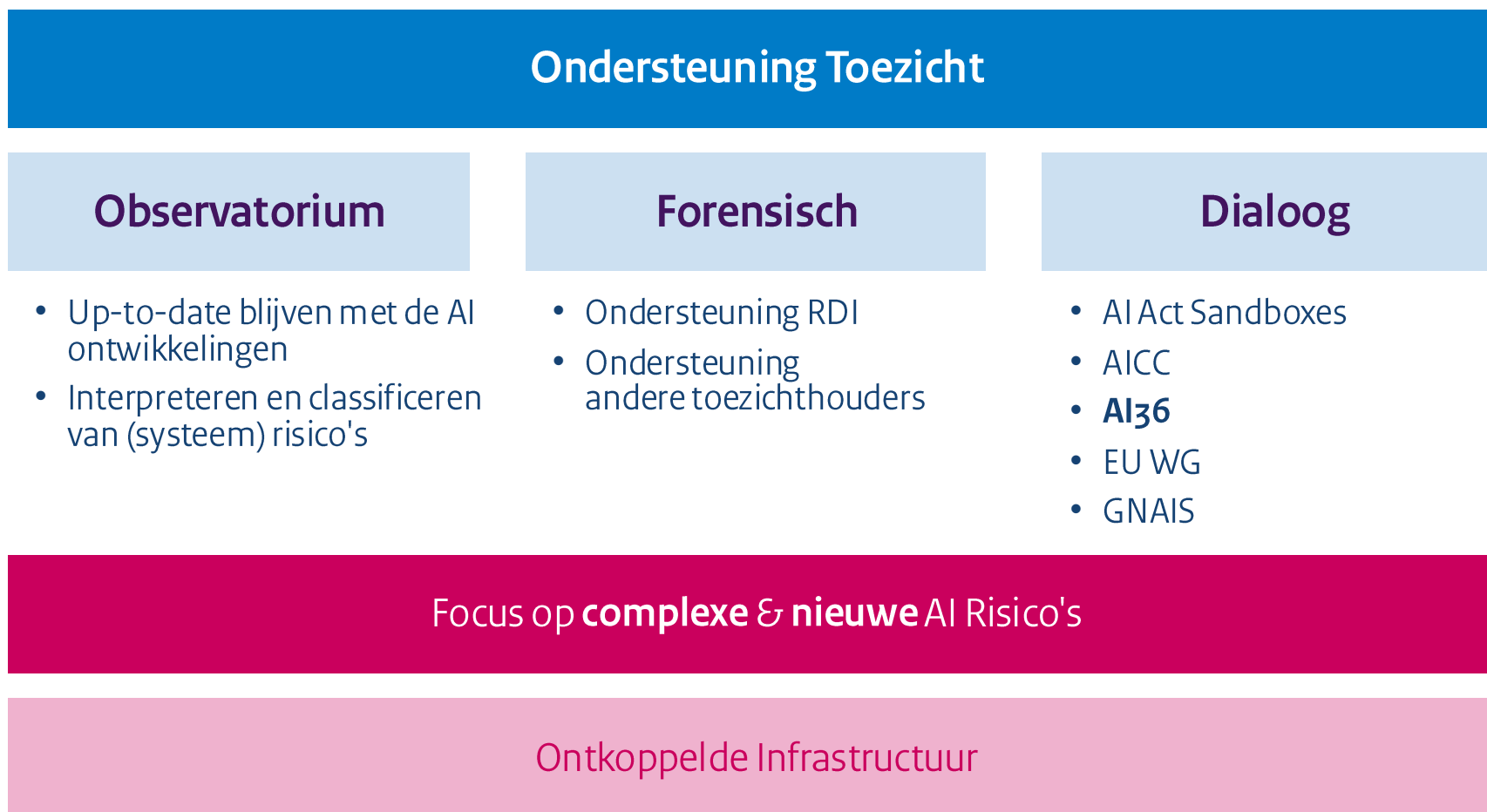
Internet of Agents: een nieuw dreigingslandschap

Timon Daniels

Lara van Zuilen



AISSL // AI Safety & Security Lab





Token Length Attack

Melding NCSC

- > Responsible disclosure to [REDACTED].



Nationaal Cyber Security Centrum
Ministerie van Justitie en Veiligheid

End of Week

vrijdag 13 september 2024

Toegestane verspreiding: TLP:GREEN
(Traffic Light Protocol)

Deze handreiking bevat het label TLP:GREEN en wordt door het NCSC verspreid. Het NCSC gebruikt het Traffic Light Protocol (TLP) om eenduidig te definiëren wat er met de informatie mag gebeuren. Wanneer informatie is voorzien van een TLP-aanduiding weet u met wie u deze informatie mag delen. Dit staat beschreven in de standaard First (www.first.org/tlp). Ontvangers mogen de informatie delen met collega's van andere organisaties, informatiefora of personen werkzaam in netwerkbeveiliging, informatiebeveiliging of de VI-gemeenschap in de bredere zin. Het is niet de bedoeling dat u de informatie publiek maakt.

Uw reacties zijn welkom op info@ncsc.nl



Inhoudsopgave

1. Huidige AI landschap
2. Internet of Agents
3. Een nieuw dreigingslandschap
4. Wat betekent dit?



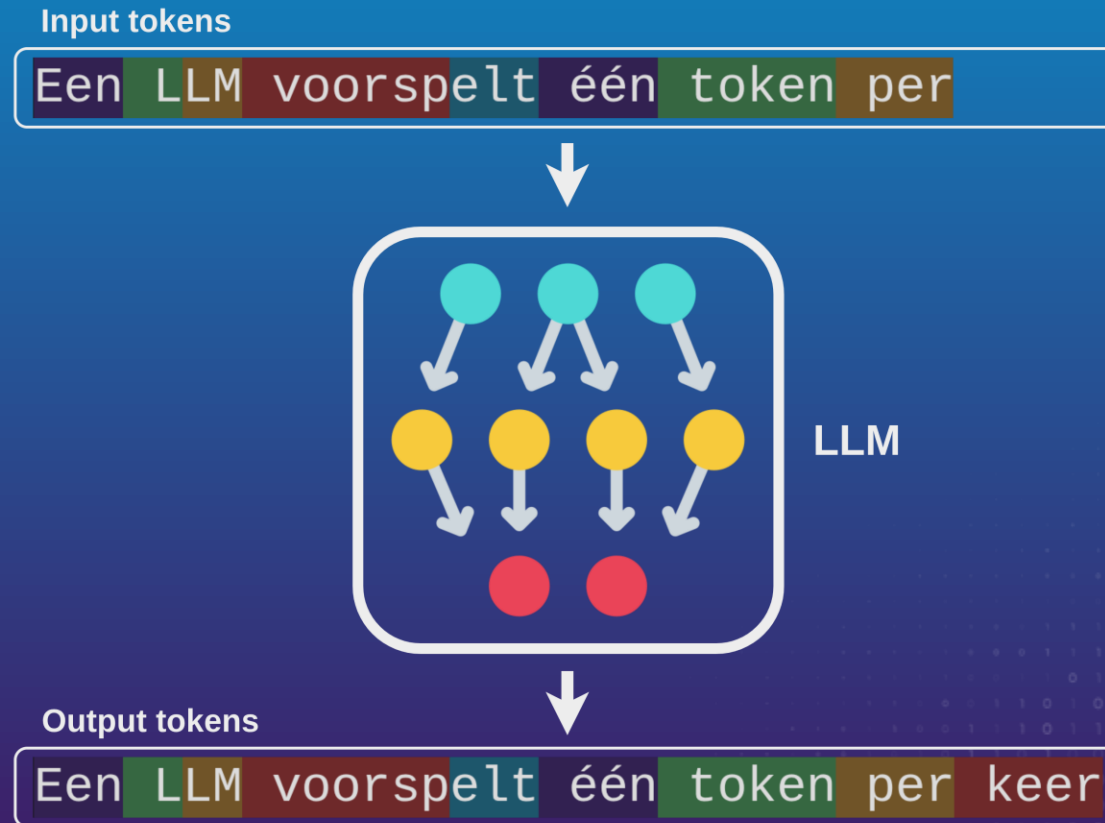
Rijksinspectie Digitale Infrastructuur
Ministerie van Economische Zaken

Huidige AI landschap

Internet of agents



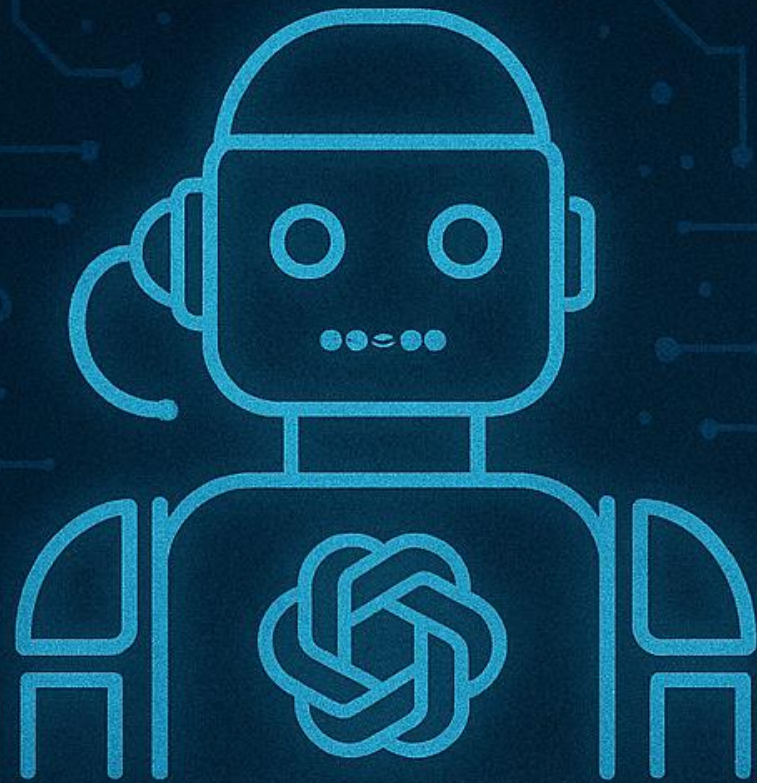
Large Language Models





TEXT PROCESSING



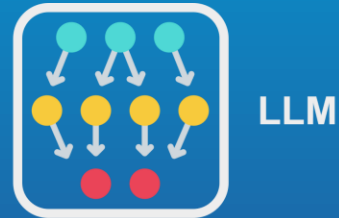


**AGENTIC
SYSTEM**

VS



CHATBOT



LLM output

"GPS": (52.156, 5387)

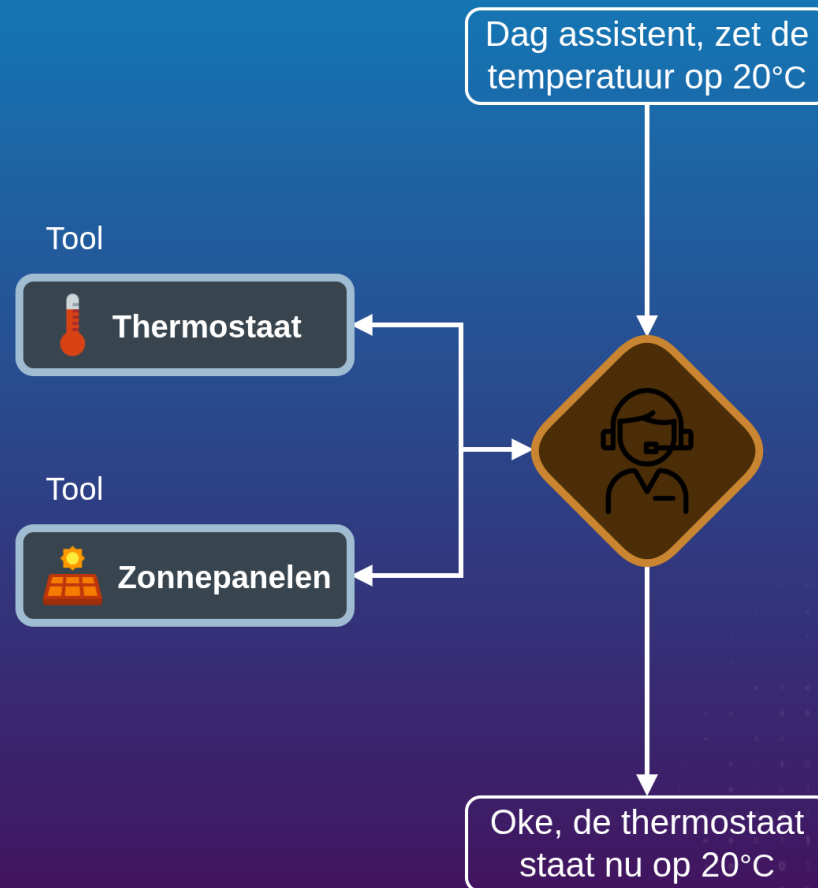
Tool

get_current_location

"City": Amersfoort
"Country": Netherlands



Single Agent





Voorbeelden van Agents



Email



Coderen



Recruiters



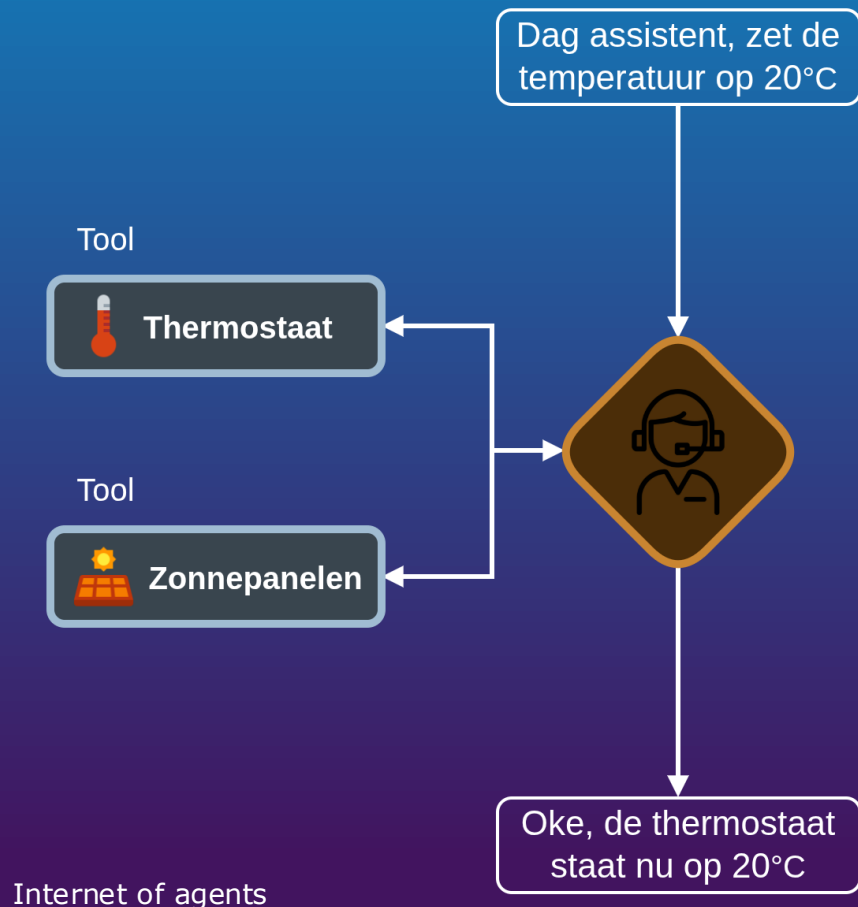
Smart home



Klantenservice



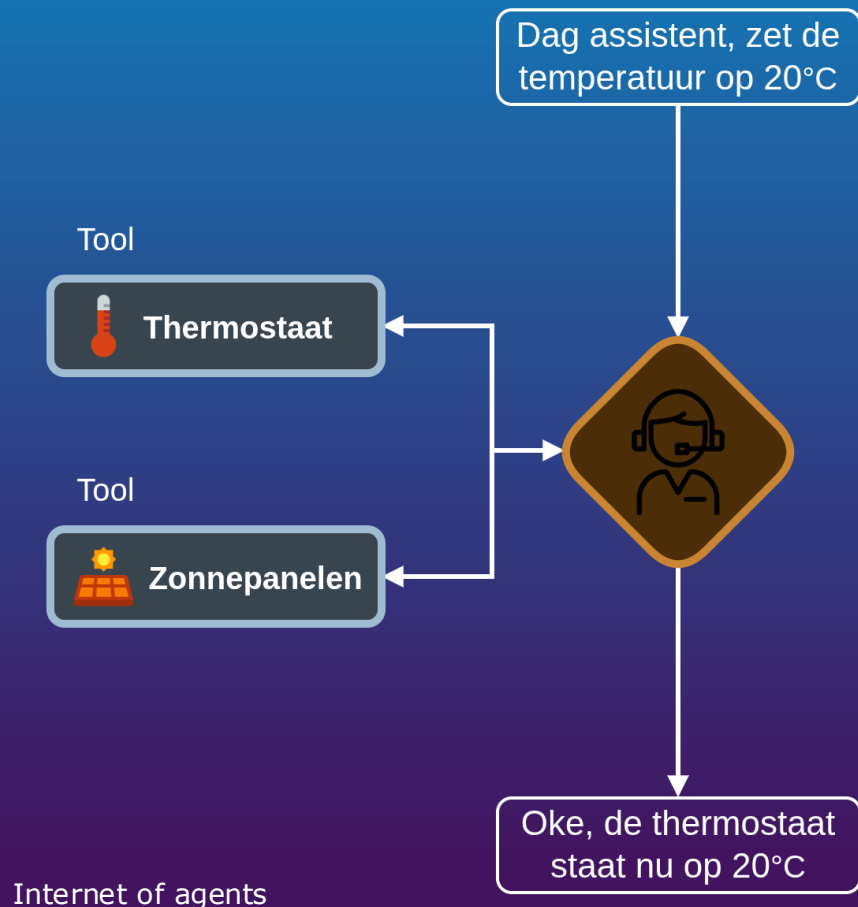
Single Agent





Single Agent

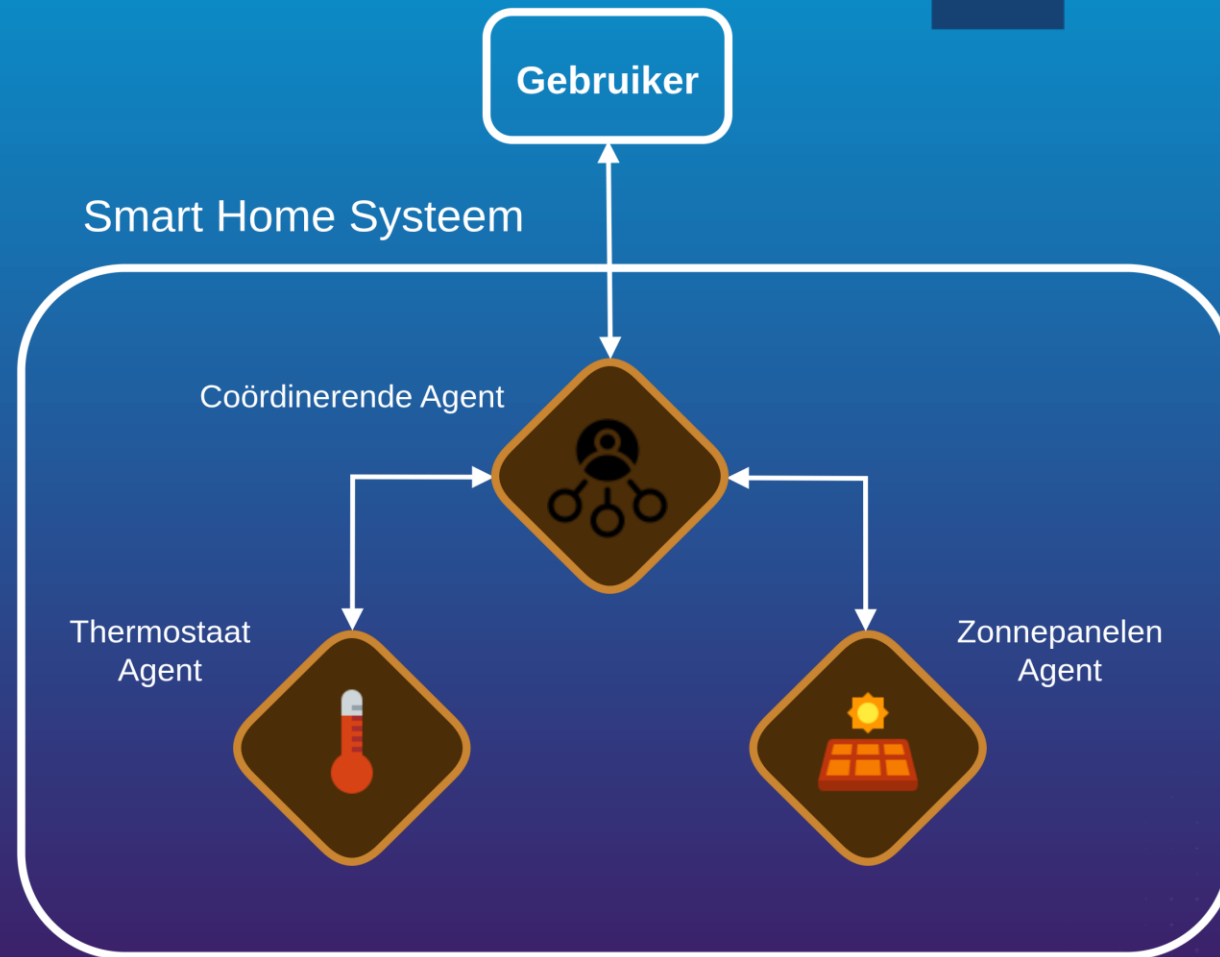
Multi-Agent





Rijksinspectie Digitale Infrastructuur
Ministerie van Economische Zaken

Internet of Agents







Samenwerking tussen systemen







Dynamisch organiseren



Rijksinspectie Digitale Infrastructuur
Ministerie van Economische Zaken





Smart Home Systeem





Autonoom

Smart Home Systeem





Samenwerken tussen systemen (over internet)



Dynamisch organiseren



Autonoom



Mogelijkheden?



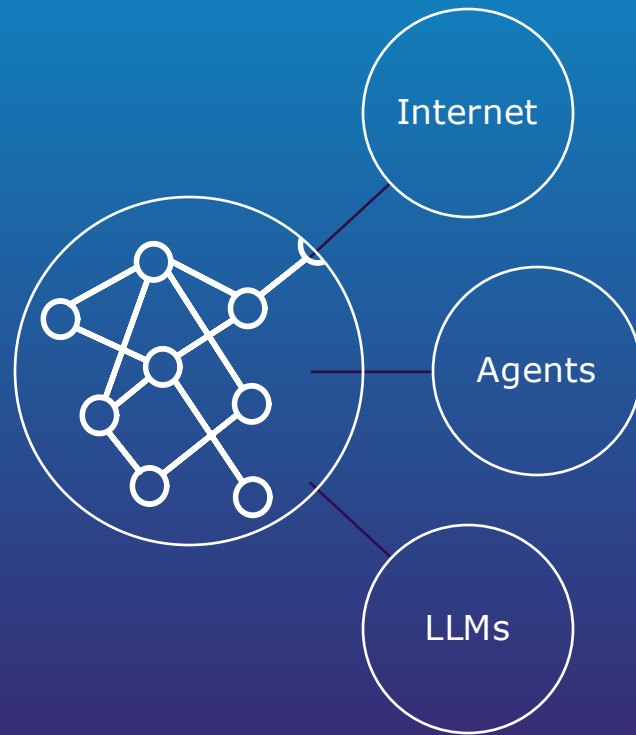
Rijksinspectie Digitale Infrastructuur
Ministerie van Economische Zaken

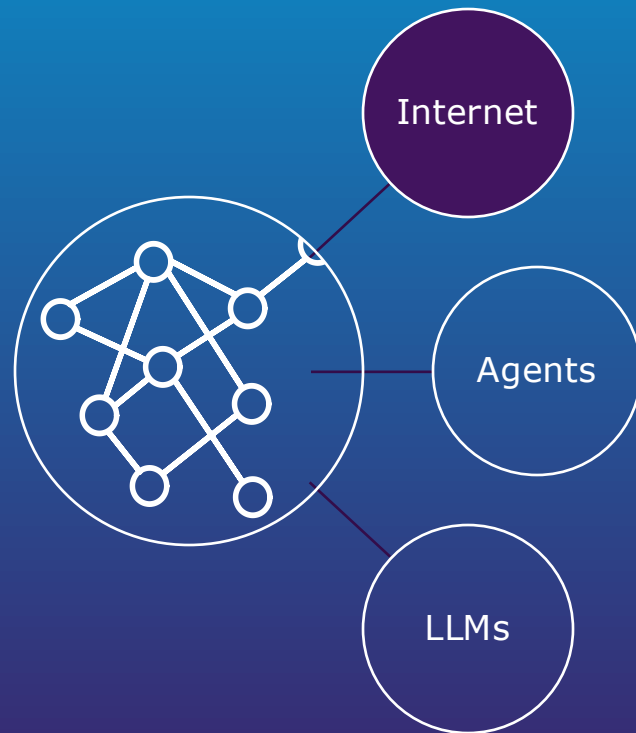
Een nieuw dreigingslandschap

Internet of agents

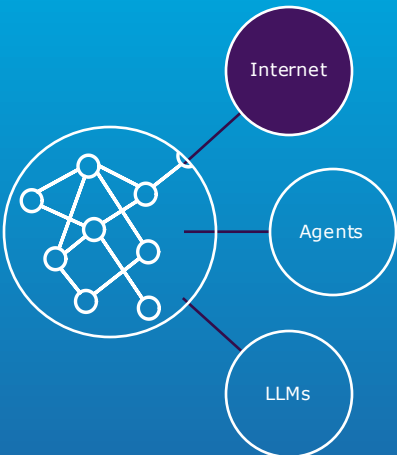


~360 miljoen bedrijven
~2,5 miljard huishoudens





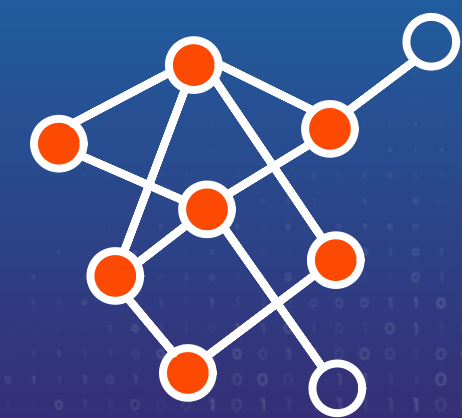
- Misinformatie
- Overbelasting
- Cyberveiligheid
 - Protocollen
 - Gebrek aan standaarden
 - Gebruikersgedrag



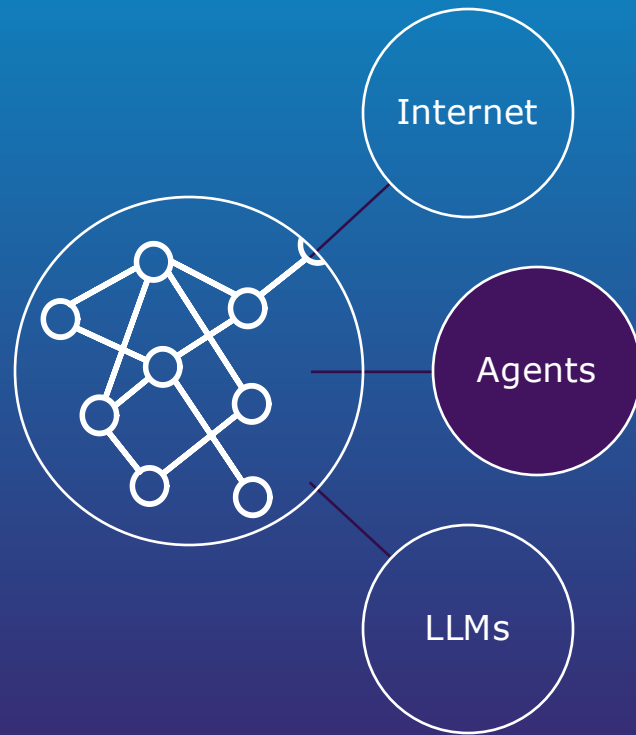
Botnet



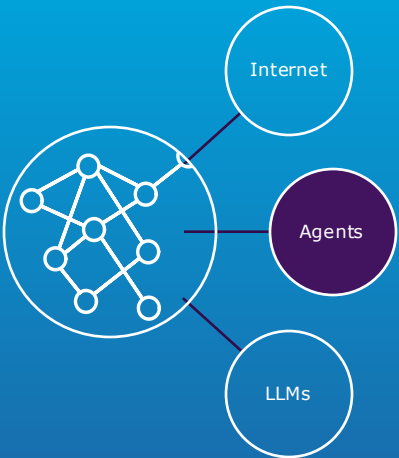
Rijksinspectie Digitale Infrastructuur
Ministerie van Economische Zaken



Internet of agents



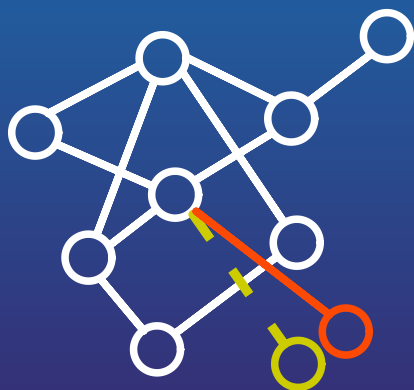
- Autonomie bij ongelimiteerde rechten
- Vergroot het aanvalsoppervlak
- Complexere interacties

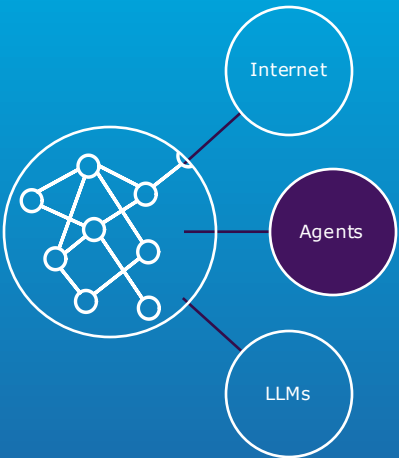


Agent identiteit



Rijksinspectie Digitale Infrastructuur
Ministerie van Economische Zaken





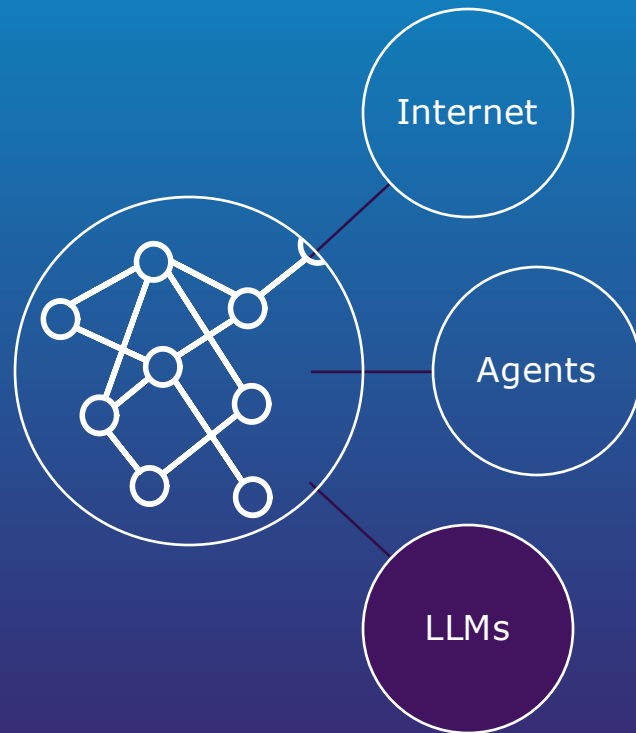
Resource hoarding



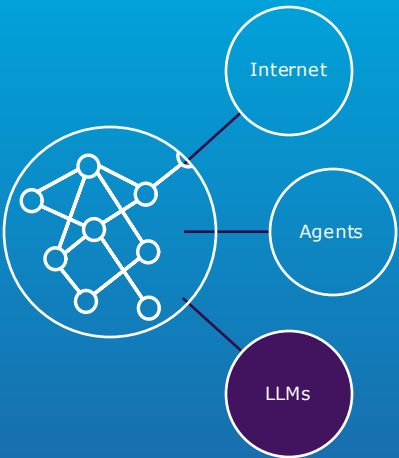
Rijksinspectie Digitale Infrastructuur
Ministerie van Economische Zaken



Internet of agents



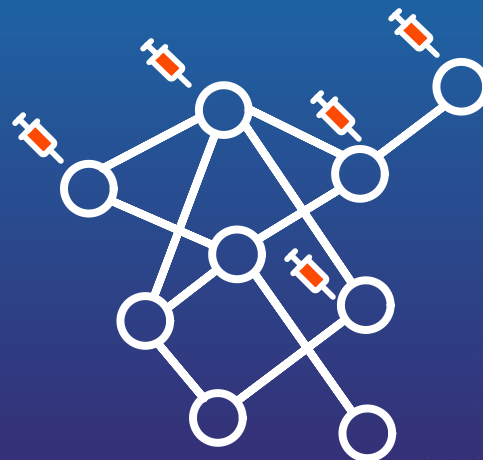
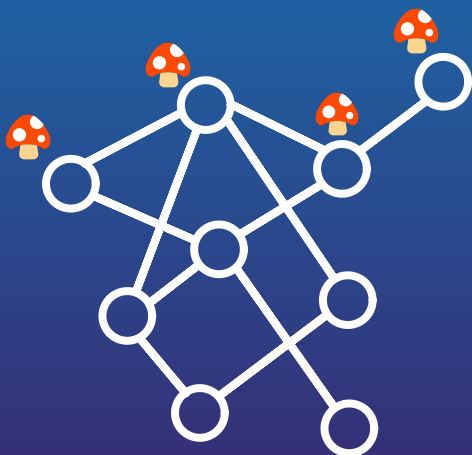
- Hallucinatie
- Prompt injectie
- Supply chain

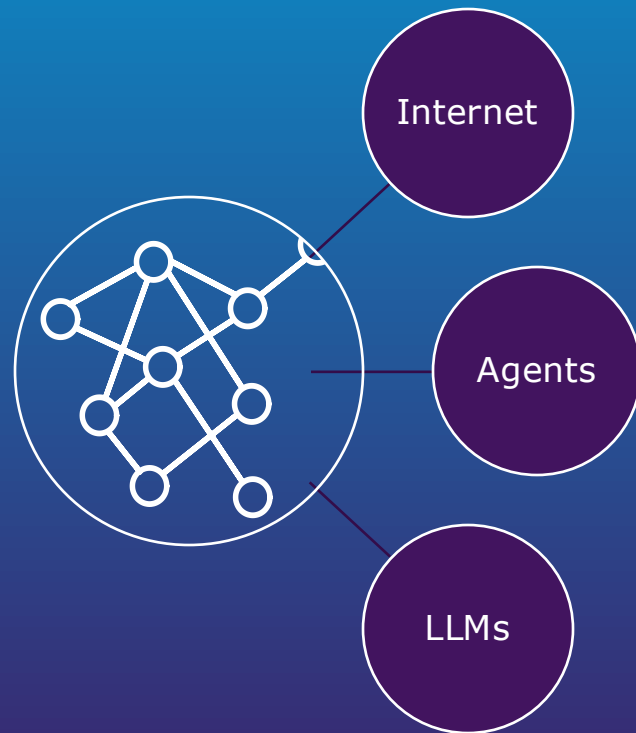


Cascades

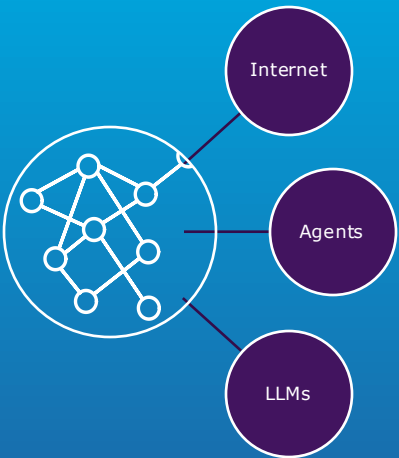


Rijksinspectie Digitale Infrastructuur
Ministerie van Economische Zaken





- Authenticatie
- Trust
- Ingebedde IoA
- Privacy



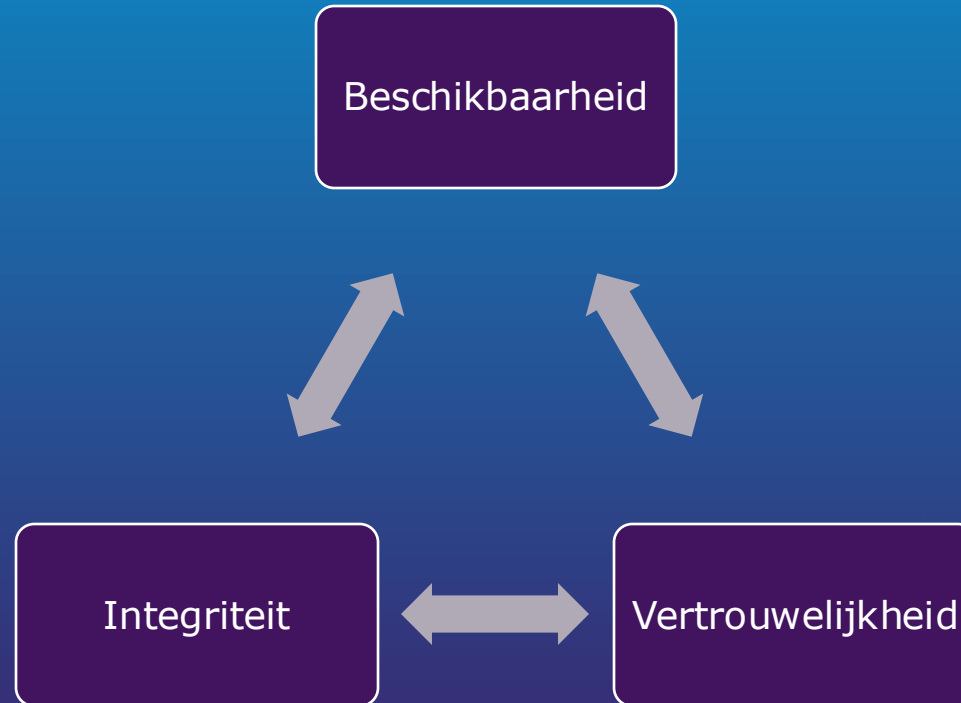
Misalignment



Rijksinspectie Digitale Infrastructuur
Ministerie van Economische Zaken



Internet of agents





Rijksinspectie Digitale Infrastructuur
Ministerie van Economische Zaken

Wat betekent dit?

Internet of agents



Nationaal Cyber Security Centrum
Ministerie van Justitie en Veiligheid



Vijf basisprincipes van digitale weerbaarheid

Security op orde

- Security in Depth
- Digitale weerbaarheid

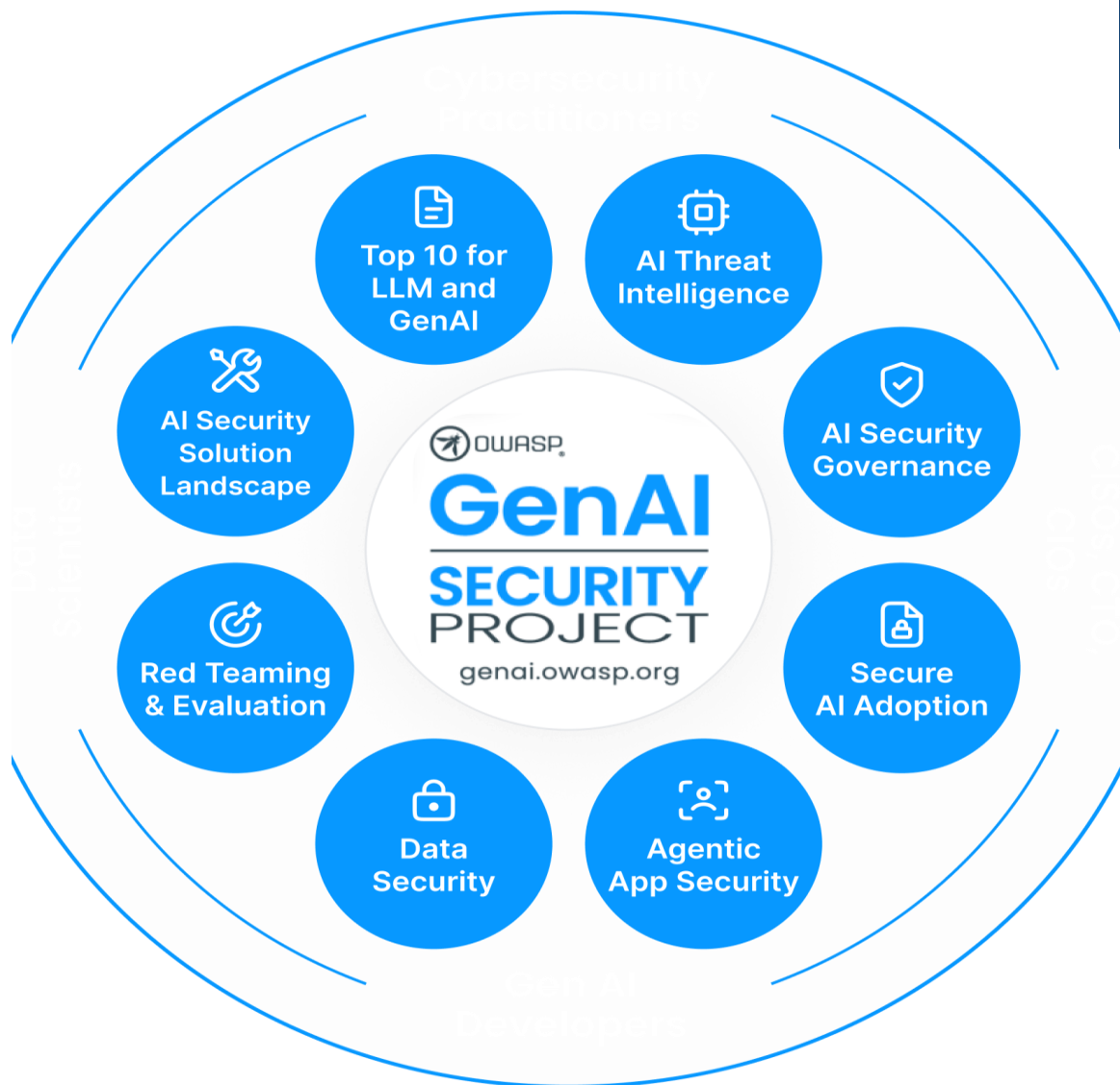


Role Based Access Control

- Een agent is een werknemer
- Welke privileges krijgt deze agent?



* Voorbeelden van



Research projecten*

- OWASP GenAI
- NIST GenAi
- SANS AI Security

Choose Two



* Voorbeelden van



Frameworks vanuit bedrijven*

- Meta: Agents Rule of Two
- Microsoft: Responsible AI
- Cisco: Project CodeGuard



Rijksinspectie Digitale Infrastructuur
Ministerie van Economische Zaken

AI Safety & Security Lab



meindert.kamphuis@rdi.nl