THE COGNITIVE HOURGLASS:
AGENT ABSTRACTIONS IN THE LARGE MODELS ERA

A. RICCI, S. BURATTINI
University of Bologna, Italy

S. MARIANI, F. ZAMBONELLI,
University of Modena-Reggio Emilia, Italy

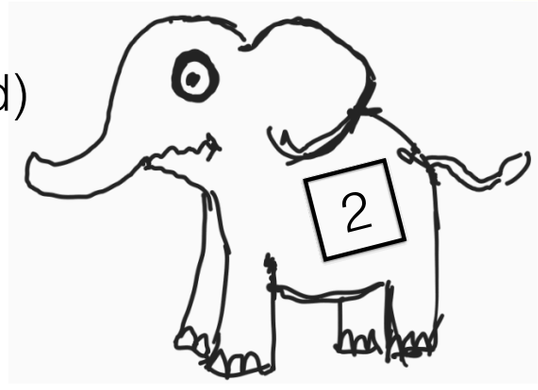
C. CASTELFRANCHI
ISTC-CNR, Italy

AAMAS BLUESKY + EMAS 2024

PREQUEL

- A Future for Agent Programming (2015)
  [B. Logan, EMAS 2015 invited talk]
- Agent Programming in the Cognitive Era (2020)
  [R. Bordini, B. Logan, K. Hindriks, A. El-Falla Segrouchni, A. Ricci - JAAMAS]
- Agent Programming in the Cognitive Era:
  A New Era for Agent Programming? (2021)
  [A. Ricci, EMAS 2021 invited talk]

(Agent-Oriented)
Programming
?

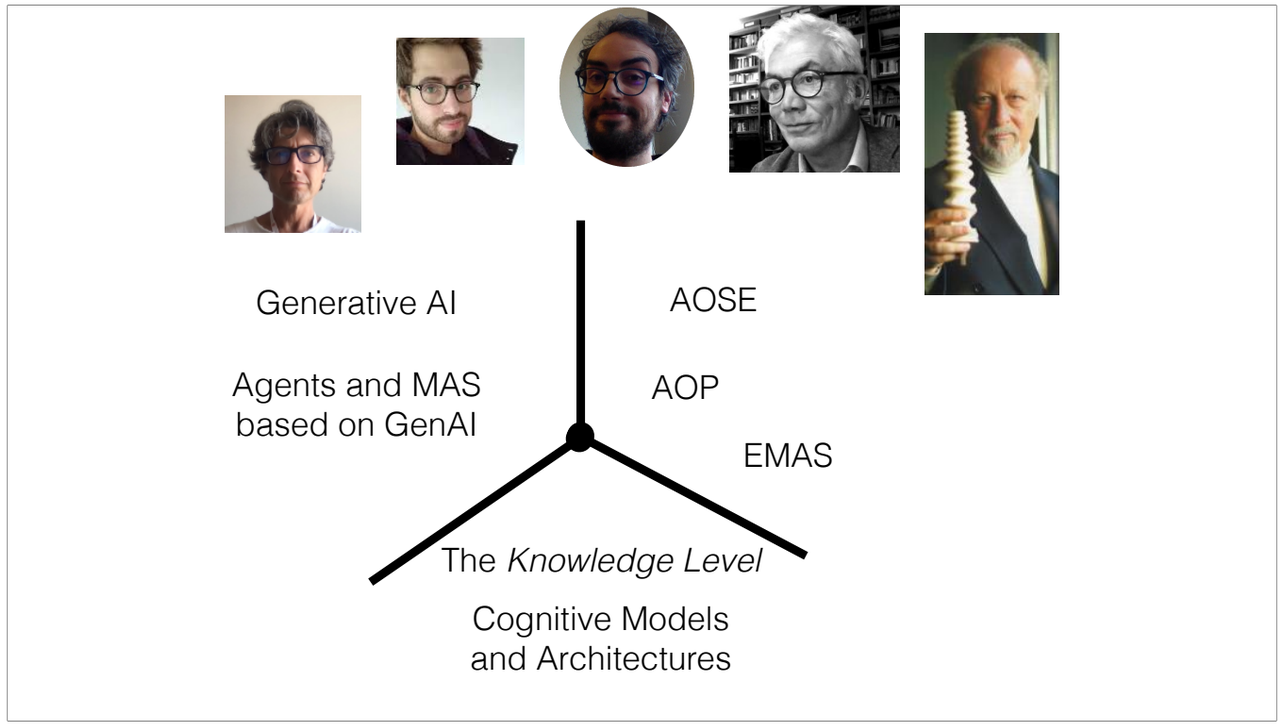


“Software 2.0”

Machine Learning

data-driven

Gen-AI


Generative AI
Agents and MAS
based on GenAI
AOSE
AOP
EMAS
The *Knowledge Level*
Cognitive Models
and Architectures

THE DANGER OF **ELIMINATIVISM**

- Deeming higher-level abstractions *unnecessary* once lower-level mechanisms are understood
  - deeming macro-level *cognitive concepts* unnecessary, relying only on micro-level implementing mechanisms

GENERAL PRINCIPLE

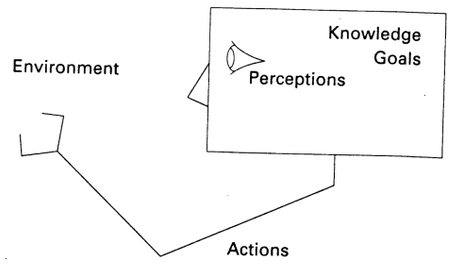
- Importance of **Abstraction** and **Levels of Abstraction**
  - both from a *scientific* and *engineering* perspective
- To specify | understand | explain | predict | control the behaviour of complex artificial/natural systems

THE KNOWLEDGE LEVEL

[Newell, 1982, 1993]

Hierarchy of computer systems

Knowledge-level systems Medium: Knowledge Laws: Principle of Rationality
Program-level systems Medium: Data structures, programs Laws: Sequential interpretation of programs
Register-transfer system Medium: Bit vectors Laws: Parallel logic
Logic circuits Medium: Bits Laws: Boolean algebra
Electric circuits Medium: Voltage/current Laws: Ohm's law, Kirchhoff's law
Electronic devices Medium: Electrons Laws: Electron physics



The knowledge-level system
(aka “an agent”)

~Dennett’s **Intentional Stance**

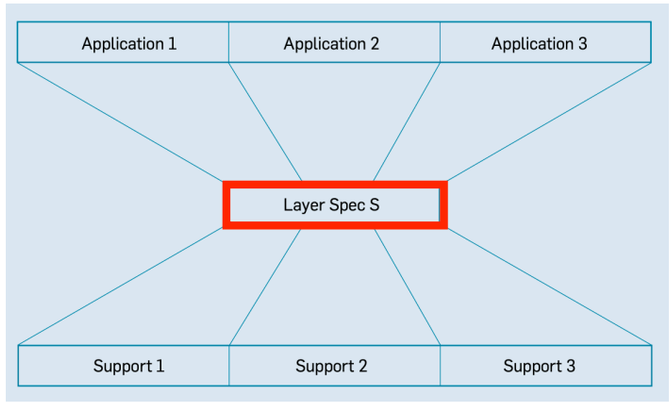
EXTENDED TO SOCIAL LEVEL

[Jennings, 2000]

Dimension	**Description**	**Knowledge level**	**Social level**
System	Entity to be described	(asocial) Agent	Agent organisation
Components	The system's primitive elements	Goals, Actions	Agents, Interaction channels, Dependencies, Organisational relationships
Compositional law	How the components are assembled	Various	Roles, Organisation's rules
Behaviour law	How the system's behaviour depends upon its composition and components	Principle of rationality	Principle of organisational rationality
Medium	The elements to be processed to obtain the desired behaviour	Knowledge	Organisation and social obligations, Means of influencing others, Means of changing organisational structures

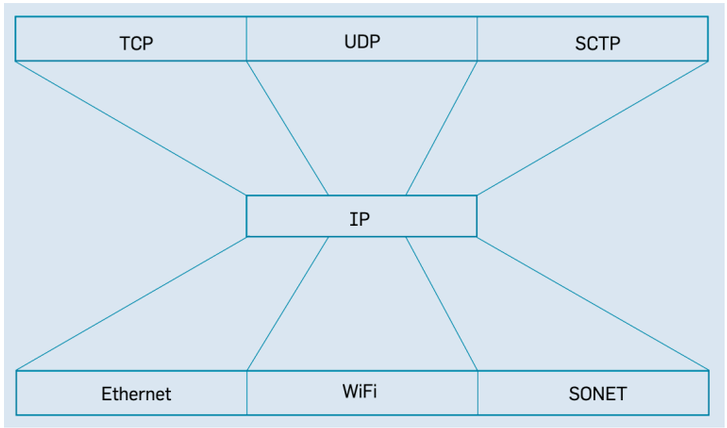
THE HOURGLASS MODEL

- *Design* blueprint
  - great diversity of applications
  - great diversity of supporting services
- Distinguished layer in the center (narrow waist or *neck*)
  - a stack of *abstractions*
    - the sole means of accessing the lower-level resources of the system

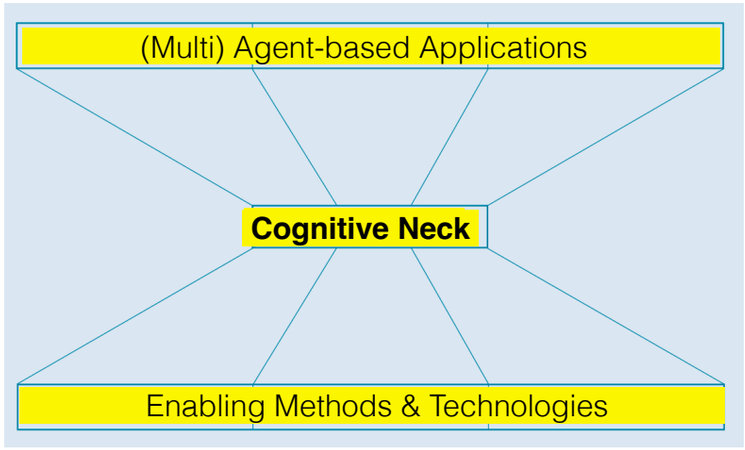


[Beck 2019, CACM]

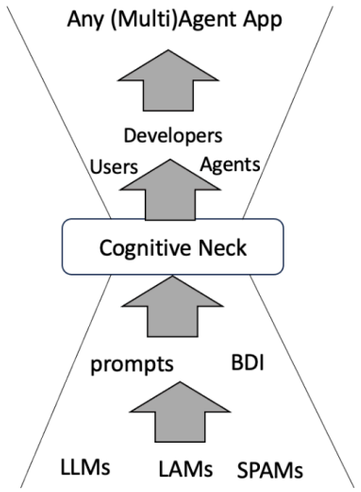
THE INTERNET HOURGLASS



A COGNITIVE HOURGLASS

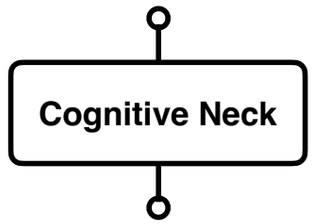


A COGNITIVE HOURGLASS

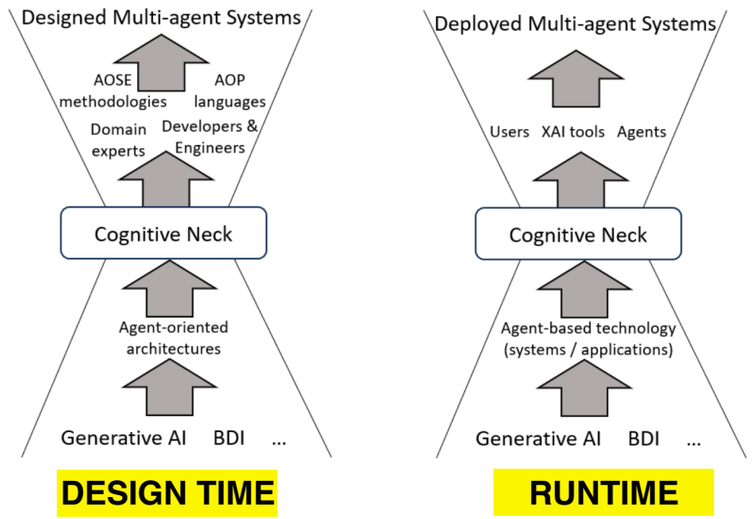


THE COGNITIVE NECK

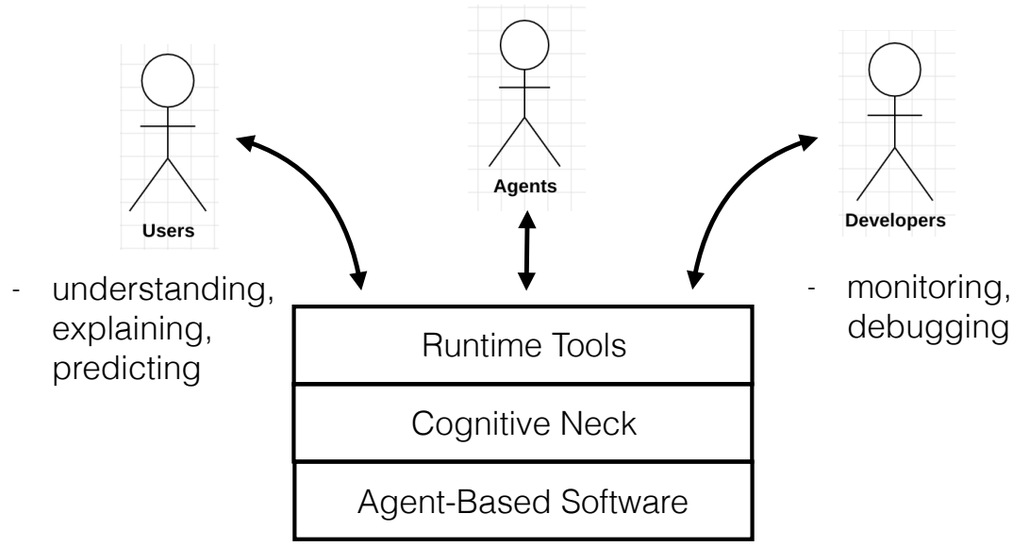
- An *abstraction barrier* to preserve the KL
- An effective enabler to exploit theories, methods & mechanisms



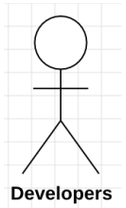
A COGNITIVE HOURGLASS



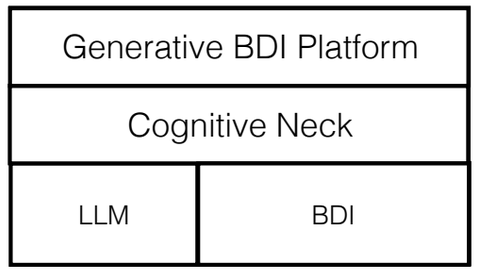
"VERTICAL" EXAMPLES



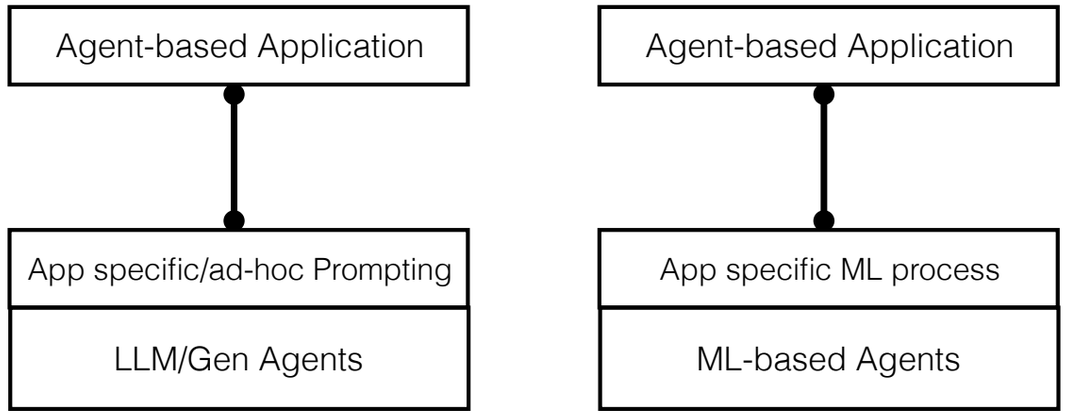
"VERTICAL" EXAMPLES



- design
- programming



AVOIDING ELIMINATIVISM



WHAT

STACK OF ABSTRACTIONS

IN THE COGNITIVE NECK?

SOME REQUIREMENTS

- To be *“human-compatible”* [Russell 2020]

  => cognitive, at the Knowledge Level
- To be effective as domain-independent meta-model

  => to be applicable to any domain
- To be effective in exploiting “AI masteries”
- To feature *learning* as developmental core capability
- To cross the full engineering processes

ONGOING | FUTURE WORK

- Working on the stack of abstractions in the neck
- Developing & refining the conceptual framework by using concrete examples and use cases
  - one is about the “Generative BDI” case, applied to specific application domains
- Exploring the value (?) of the framework to rethink & refine the full development process and related tools

<connection with previous research about learning>

THANKS.