Symbolic AI ??? What It Can and Cannot Do

Thomas Eiter

Institute of Logic and Computation
Knowledge-Based Systems Group

Vienna University of Technology (TU Wien)

Workshop “Paradigm Shift in Computer Science??”
TU Wien, Vienna, November 28, 2024

Influential Ideas

Symbolic Artificial Intelligence is the term for the collection of all methods in artificial intelligence research that are based on high-level "symbolic" (human-readable) representations of problems, logic and search. (Marvin Minsky)

One path to human-level AI uses mathematical logic to formalise common-sense knowledge in such a way that common-sense problems can be solved by logical reasoning. (John McCarthy)

Influential Ideas

Symbolic Artificial Intelligence is the term for the collection of all methods in artificial intelligence research that are based on high-level "symbolic" (human-readable) representations of problems, logic and search. (Marvin Minsky)

One path to human-level AI uses mathematical logic to formalise common-sense knowledge in such a way that common-sense problems can be solved by logical reasoning. (John McCarthy)

- Newell and Simon's symbol processing hypothesis (1976)
  A physical symbol system has the necessary and sufficient means for general intelligent action.

Influential Ideas

Symbolic Artificial Intelligence is the term for the collection of all methods in artificial intelligence research that are based on high-level "symbolic" (human-readable) representations of problems, logic and search. (Marvin Minsky)

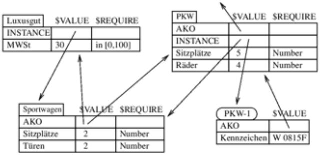
One path to human-level AI uses mathematical logic to formalise common-sense knowledge in such a way that common-sense problems can be solved by logical reasoning. (John McCarthy)

- Newell and Simon's symbol processing hypothesis (1976)
  A physical symbol system has the necessary and sufficient means for general intelligent action.
- Lenat and Feigenbaum's knowledge principle (1987)
  A system exhibits intelligent understanding and action at a high level of competence primarily because of the specific knowledge that it can bring to bear: the concepts, facts, representations, methods, models, metaphors, and heuristics about its domain of endeavor.

Influential Ideas, cont'd

- Minsky's frames (1975)

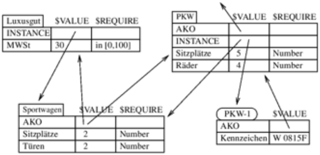
When one encounters a new situation (or makes a substantial change in one's view of the present problem) one selects from memory a substantial structure called a frame. This is a remembered framework to be adapted to fit reality by changing details as necessary ...



Influential Ideas, cont'd

- Minsky's frames (1975)

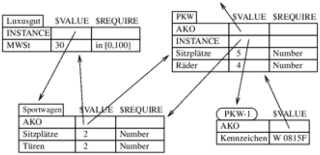
  When one encounters a new situation (or makes a substantial change in one's view of the present problem) one selects from memory a substantial structure called a frame. This is a remembered framework to be adapted to fit reality by changing details as necessary . . .



- Conceptual graphs (Sowa, 1976,1984)
  - SNePS (Shapiro, 1979) , NETL (Fahlmann, 1979)
  - KL-ONE (Brachman et al., 1979) and relatives (CLASSIC BACK, KRIS)
  - emergence of description logics

Influential Ideas, cont'd

- Minsky's frames (1975)



When one encounters a new situation (or makes a substantial change in one's view of the present problem) one selects from memory a substantial structure called a frame. This is a remembered framework to be adapted to fit reality by changing details as necessary ...

- Conceptual graphs (Sowa, 1976,1984)
  - SNePS (Shapiro, 1979) , NETL (Fahlmann, 1979)
  - KL-ONE (Brachman et al., 1979) and relatives (CLASSIC BACK, KRIS)
  - emergence of description logics

- Shoham's agent-oriented programming (1993)
  - computational framework based on *societal computation and interaction*
  - key idea: program agents directly by notions like beliefs, intentions, and goals
  - use intentional stance to program machines
  - refine & enriche object oriented programming

What Early Symbolic Systems Could Do

- Playing Games
  - checkers, chess (but beating top humans took time)

What Early Symbolic Systems Could Do

- Playing Games
  - checkers, chess (but beating top humans took time)
- Theorem proving
  - limited scope, but traits of generative AI

What Early Symbolic Systems Could Do

- Playing Games
  - checkers, chess (but beating top humans took time)
- Theorem proving
  - limited scope, but traits of generative AI
- Simple conversations & natural language understanding
  - Weizenbaum's ELIZA (1968)
  - Winograd's SHRDLU ($\sim$1969)

What Early Symbolic Systems Could Do

- Playing Games
  - checkers, chess (but beating top humans took time)
- Theorem proving
  - limited scope, but traits of generative AI
- Simple conversations & natural language understanding
  - Weizenbaum's ELIZA (1968)
  - Winograd's SHRDLU ($\sim$1969)
- Solving planning problems
  - landmark: STRIPS

What Early Symbolic Systems Could Do

- Playing Games
  - checkers, chess (but beating top humans took time)
- Theorem proving
  - limited scope, but traits of generative AI
- Simple conversations & natural language understanding
  - Weizenbaum's ELIZA (1968)
  - Winograd's SHRDLU ($\sim$1969)
- Solving planning problems
  - landmark: STRIPS

Shakey the robot (SRI)

What Early Symbolic Systems Could Do

- Playing Games
  - checkers, chess (but beating top humans took time)
- Theorem proving
  - limited scope, but traits of generative AI
- Simple conversations & natural language understanding
  - Weizenbaum's ELIZA (1968)
  - Winograd's SHRDLU ($\sim$1969)
- Solving planning problems
  - landmark: STRIPS
- Mastering domain-specific problems
  - "expert systems" for medical diagnosis, configuration, ore deposit assessment . . .

Shakey the robot (SRI)

What Early Symbolic Systems Could Do

- Playing Games
  - checkers, chess (but beating top humans took time)
- Theorem proving
  - limited scope, but traits of generative AI
- Simple conversations & natural language understanding
  - Weizenbaum's ELIZA (1968)
  - Winograd's SHRDLU (∼1969)
- Solving planning problems
  - landmark: STRIPS
- Mastering domain-specific problems
  - "expert systems" for medical diagnosis, configuration, ore deposit assessment . . .

Shakey the robot (SRI)

Use of search, logic, probabilities/uncertainty measures, hybrid methods

What Early Symbolic AI Could Not Do (Keeps Struggling)

- Vision
  - more general, sensory input
- Deeper Natural Language Understanding
- Deal with exploding search spaces
  - NP-hardness: the kiss of death
- Go beyond limited tasks
  - “narrow AI”

What Early Symbolic AI Could Not Do (Keeps Struggling)

- Vision
  - more general, sensory input
- Deeper Natural Language Understanding
- Deal with exploding search spaces
  - NP-hardness: the kiss of death
- Go beyond limited tasks
  - “narrow AI”

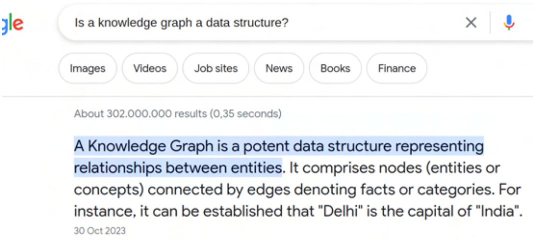
Remark: Alan Turing (1950)

- suggested major components of AI:
  knowledge, reasoning, language understanding, learning
- build a learning machine and teach it

Advances of Symbolic AI

- Semantic Systems
  (esp. Knowledge Graphs)
- Games
- Solving Mathematical Problems
- Verfication
- Combinatorial Optimisation
- Configuration
- Planning, Scheduling
- Declarative Programming
- Multi-Agent Systems
- ...

Querying the Web



- The usefulness of the Web hinges on the idea of adding semantics
- Symbolic knowledge representation and reasoning are at the core
  - Web Ontology language for semantic markup
- Google's knowledge graph (2016) is the backbone of semantic reasoning
  - revival of conceptual graphs
- Beyond, large scale conceptual reasoning (e.g. SNOMED system)

Games

- Chinook (checkers, 1994)
- Deep Blue (chess, 1997)
- Libratus/Pluribus (poker, 2017/2019)

  Heads-up, No-Limit Texas Hold’em

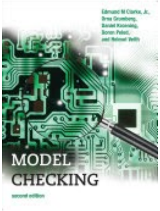
  - highly complex game: 10^{160} play paths
  - breakthrough on strategic reasoning with imperfect information: analyse *own weaknesses*, not only the opponent’s
  - Pluribus (multiple players) needs no super-computer

Solving Mathematical Problems

- 4-Colour Theorem (1974)
  - historic example
- Checkers solved (2007)
  - not only heuristics but certainty
- Kepler's Conjecture (2017)
  - use of proof assistants (HOL Light, Isabelle)
- Pythagorean Triple Problem (2024)
  - 200 Terabyte of space, logic and constraint techniques
- ...

Verification

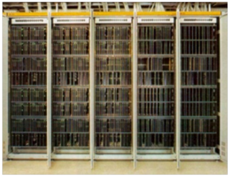
- Undecidable problem
- Landmark: Intel's Pentium FDIV bug (Clarke et al., 1996)
- Symbolic Model Checking
- Proving correctness of specifications
  - fueled by enormous advances in SAT solving & automated reasoning
  - key: exploit structure
- Industrial use, by major companies
- Software industry (Amazon WS, Microsoft, . . .)
- Big potential for security
- Logical Synthesis (correctness by design)

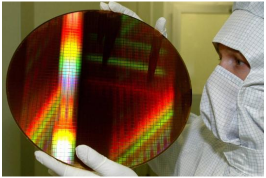


Combinatoral Optimization

Configuration and Scheduling

© Siemens AG

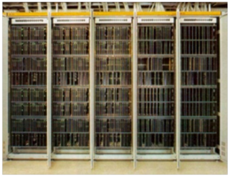


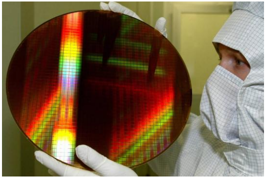


- Large scale configuration problems
  - hardware, software
  - plants
  - user interfaces, . . .
- Scheduling as temporal configuration (loosely)
  - job processing
  - tournaments . . .

Configuration and Scheduling

© Siemens AG





- Large scale configuration problems
  - hardware, software
  - plants
  - user interfaces, . . .
- Scheduling as temporal configuration (loosely)
  - job processing
  - tournaments . . .
- Learning of limited value:
  - ad-hoc formulation, hard constraints, missing data, changes

What's AI Missing?

- From a **cognitive perspective**, a lot:
  - understanding of the mind
    "brains cause minds" (J. Searle), but how?
  - deeper real world reasoning
  - abstraction
  - commonsense
    - objects fall down by gravity, a selfish person will not share its cookie 'behind' is transitive, . . .

abstraction
reasoning
learning
perception

What's AI Missing?

- From a **cognitive perspective**, a lot:
  - understanding of the mind
    "brains cause minds" (J. Searle), but how?
  - deeper real world reasoning
  - abstraction
  - commonsense
    - objects fall down by gravity, a selfish person will not share its cookie 'behind' is transitive, ...

abstraction
reasoning
learning
perception

- From a **technological perspective**, too:
  - scalability and efficiency
  - robustness
  - validation and verifiability
  - explainability: what and why $\rightarrow$ causality

What's AI Missing?

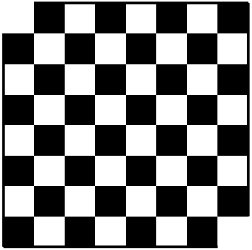
- From a **cognitive perspective**, a lot:
  - understanding of the mind
    "brains cause minds" (J. Searle), but how?
  - deeper real world reasoning
  - abstraction
  - commonsense
    - objects fall down by gravity, a selfish person will not share its cookie 'behind' is transitive, ...

abstraction
reasoning
learning
perception

- From a **technological perspective**, too:
  - scalability and efficiency
  - robustness
  - validation and verifiability
  - explainability: what and why $\rightarrow$ causality
- As well as from a **social perspective**:
  - human / AI interaction
  - fairness and unbiased data
  - norms and ethics

Issue: Abstraction

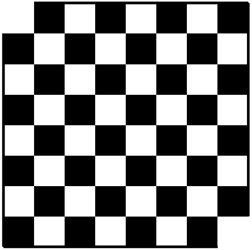
Mutilated Chessboard Problem:

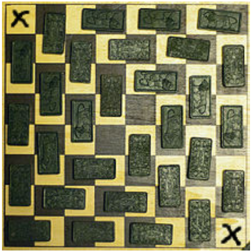


- suppose a chessboard has two diagonally opposite corners removed
- *Is it possible to place dominoes of size 2x1 so as to cover all remaining squares?*

Issue: Abstraction

Mutilated Chessboard Problem:

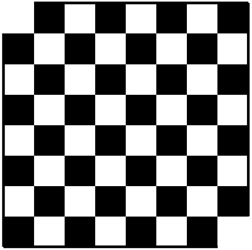


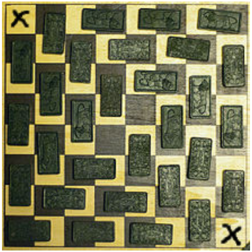


- suppose a chessboard has two diagonally opposite corners removed
- *Is it possible to place dominoes of size 2x1 so as to cover all remaining squares?*

Issue: Abstraction

Mutilated Chessboard Problem:

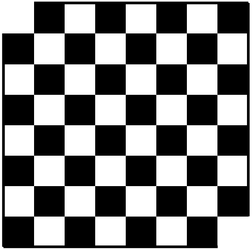


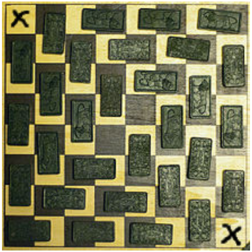


- suppose a chessboard has two diagonally opposite corners removed
- *Is it possible to place dominoes of size 2x1 so as to cover all remaining squares?*
- ChatGPT: gives the right answer, with a sound explanation

Issue: Abstraction

Mutilated Chessboard Problem:





- suppose a chessboard has two diagonally opposite corners removed
- *Is it possible to place dominoes of size 2x1 so as to cover all remaining squares?*
- ChatGPT: gives the right answer, with a sound explanation
- but for the follow up: *"If the board has a size different from 8x8, is it then possible?"*, it says "yes" and gives 6x6 as example...

Issue: Problem Solving

Well it's about traveling. A car is traveling from a city A to a city B at a distance of 50 km at a speed of 30 km/h. Another car, a blue volvo, starts 10 minutes later and travels the same route , but with 60 km/h. How far is that car going?

This is a classic relative motion problem. Let's break it down step by step.

Known facts:

1. **Car A**:
   - Distance between cities A and B: 50 km
   - Speed: $30\,\mathrm{km/h}$
2. **Car B (Blue Volvo)**:
   - Starts 10 minutes ($\frac{1}{6}$ hour) later
   - Speed: $60\,\mathrm{km/h}$

We are tasked to find how far **Car B** will travel before catching up with **Car A**.

- Listen (understanding questions) "where is the trick"
- Still fails on more involved problems (puzzles, planning etc.)

Need for Symbolic AI

Hypothesis: Symbolic AI techniques are instrumental for

Need for Symbolic AI

Hypothesis: Symbolic AI techniques are instrumental for

- Deeper World Reasoning
  - Causality

Need for Symbolic AI

Hypothesis: Symbolic AI techniques are instrumental for

- Deeper World Reasoning
  - Causality
- Commonsense

Need for Symbolic AI

Hypothesis: Symbolic AI techniques are instrumental for

- Deeper World Reasoning
  - Causality
- Commonsense
- Verified systems
  - specfication needs formal language

Need for Symbolic AI

Hypothesis: Symbolic AI techniques are instrumental for

- Deeper World Reasoning
  - Causality
- Commonsense
- Verified systems
  - specfication needs formal language
- Ethical Systems
  - Norms and Rules

Need for Symbolic AI

Hypothesis: Symbolic AI techniques are instrumental for

- Deeper World Reasoning
  - Causality
- Commonsense
- Verified systems
  - specfication needs formal language
- Ethical Systems
  - Norms and Rules
- Modeling Epistemic and Mental States (agents)
  - reflection, meta-knowledge, meta-reasoning

Need for Symbolic AI

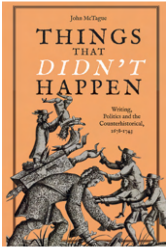
Hypothesis: Symbolic AI techniques are instrumental for

- Deeper World Reasoning
  - Causality
- Commonsense
- Verified systems
  - specfication needs formal language
- Ethical Systems
  - Norms and Rules
- Modeling Epistemic and Mental States (agents)
  - reflection, meta-knowledge, meta-reasoning
- Explainability
  - need a “model” of a system
  - caters for structure, some modularity
  - need to reason from/about the model

Need for Symbolic AI

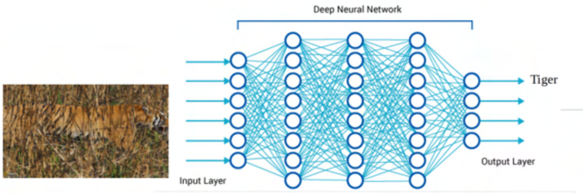
Hypothesis: Symbolic AI techniques are instrumental for

- Deeper World Reasoning
  - Causality
- Commonsense
- Verified systems
  - specfication needs formal language
- Ethical Systems
  - Norms and Rules
- Modeling Epistemic and Mental States (agents)
  - reflection, meta-knowledge, meta-reasoning
- Explainability
  - need a “model” of a system
  - caters for structure, some modularity
  - need to reason from/about the model



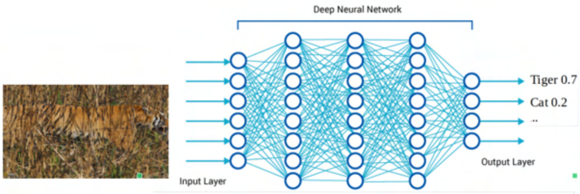
What Is Explainable AI (XAI) About?

Example: object recognition (one-hot)



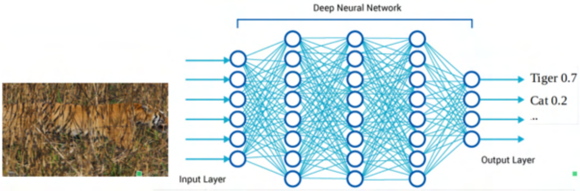
What Is Explainable AI (XAI) About?

Example: object recognition multiple (uncertainty)

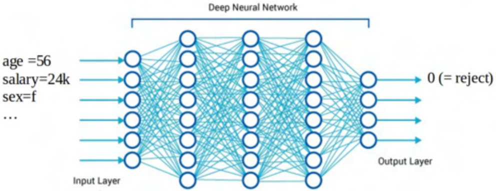


What Is Explainable AI (XAI) About?

Example: object recognition multiple (uncertainty)

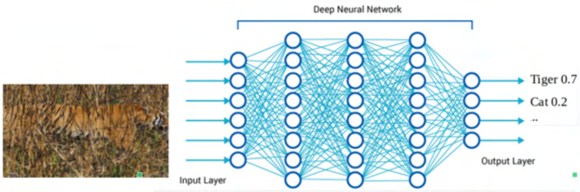


Example: mortgage application

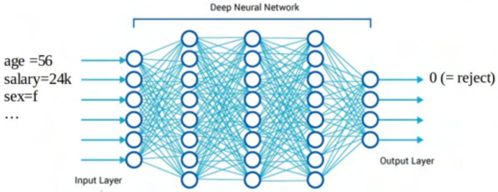


What Is Explainable AI (XAI) About?

Example: object recognition multiple (uncertainty)



Example: mortgage application



Question: Why? How comes? **Important** by EU GDPR, AI Act, etc.

Why Explainability Matters

Building Trust:

- Trust: Provides users with confidence in model outputs, especially in critical fields like finance and healthcare.
- Transparency: Helps stakeholders understand *how* decisions are reached.
- Fairness: Identifies biases or disparities, especially in regulated industries (e.g., housing, hiring).

TRUST

that we build

Example in Municipal Utilities:

Transparency in predictive models can build user confidence in areas such as resource optimization or billing automation.

Why Explainability Matters

Building Trust:

- Trust: Provides users with confidence in model outputs, especially in critical fields like finance and healthcare.
- Transparency: Helps stakeholders understand *how* decisions are reached.
- Fairness: Identifies biases or disparities, especially in regulated industries (e.g., housing, hiring).

Example in Municipal Utilities:

Transparency in predictive models can build user confidence in areas such as resource optimization or billing automation.

Symbolic AI Techniques

Currently XAI is lacking

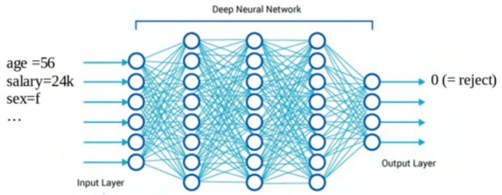
- Formal frameworks
- Warranted behaviour
- Guarantees

Symbolic AI, based on logic and formal methods, can help

General Methods and Techniques:

- Abductive reasoning as a base for explanations
- Axiom pinpointing, justification
- Formal argumentation
- ...

Logical Explainability for Classifiers



- A host of techniques (LIME, SHAP, Attention Maps, ...)
- Logic-based approach:
  - Use formulas with *feature atoms* $x_i = c$ (feature f_i has value c) resp. $x_i \geq c$ (f_i has value at least c) etc. to describe a dataset D

$$age \geq 75 \Rightarrow reject, \qquad age \leq 50 \wedge salary \geq 50k \Rightarrow accept$$
$$\ldots$$

  - Build a logical theory $T(D)$ describing the dataset D
  - Encode a neural network in this way:
    - SAT (propositional logic); MILP (mixed integer linear programming); SMT (fragments of first-order logic); ASP (answer set programming), etc.

Logical Explainability for Classifiers, cont'd

Benefits

- Exploit concepts, algorithms, and tools from logic
- A range of possible forms of explanations
  - factual explanations
  - derivations / proofs
- Aid in understanding the reasoning behind specific decisions
  - helpful for finding errors, debugging, repair
- Amenable to *reasoning* about explanations

Strong Points of Symbolic AI

- Correctness
  - soundness, completeness
- Transparency
  - inherent by design
- Transferability
  - includes abstraction (predicate languages)
- Reasoning
  - about conceptual models and their properties
  - settings of epistemic and mental states (modal logics)
  - counterfactual, nonmonotonic, and causal inferences
- Tools and Methods: rich landscape of
  - solvers (SAT, CP, SMT, ASP,. . .), highly engineered
  - calculi
  - reasoning engines, proof assistants for analysis

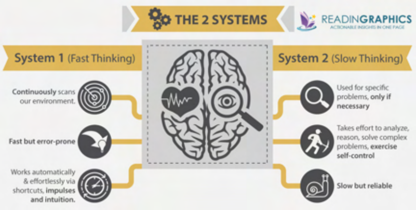
    e.g., inconsistency in Gödel's ontological proof of god (Benzmüller and Woltzenlogel, 2013)

Issues of Symbolic AI

- **Computational Cost** – still

  cf. Kahneman's Thinking, Fast and Slow (2012): processing in System 2 is much more involving than in System 1

  dealing with quantities / uncertainty



*

- **Conceptualization**
  - form the language, construct knowledge bases
- **Interfacing** human ⇔ machine
- Coping with **irrational / illogical behaviors**
  - humans are not ideal reasoners
  - cognition, psychology
- Needs **skill and expertise**

* Thinking Fast And Slow - How Good Judgement Leads To Better Decisions. Creator: Stephen Warrilow
https://readingraphics.com/book-summary-thinking-fast-and-slow/
https://i0.wp.com/readingraphics.com/uploads/2016/11/Thinking-Fast-and-Slow_the-2-systems.png

Future Development of AI

- AI View: Strong vs. weak AI
  - planes fly, submarines dive
  - still, there are models behind that are well understood
  - need reasoning from / about the models
  - learning is key to theory formation

Future Development of AI

- AI View: Strong vs. weak AI
  - planes fly, submarines dive
  - still, there are models behind that are well understood
  - need reasoning from / about the models
  - learning is key to theory formation
- Aim at a broad AI
  - beyond current narrow AI
  - neither symbolic AI nor sub-symbolic AI will suffice
  - current LLMs may be a dead end (cf. LeCun, AAAI 2024)

Future Development of AI

- AI View: Strong vs. weak AI
  - planes fly, submarines dive
  - still, there are models behind that are well understood
  - need reasoning from / about the models
  - learning is key to theory formation
- Aim at a broad AI
  - beyond current narrow AI
  - neither symbolic AI nor sub-symbolic AI will suffice
  - current LLMs may be a dead end (cf. LeCun, AAAI 2024)
- Bridging symbolic and subsymbolic AI
  - need for mental faculties is acknowledged
  - ways to achieve diverge
  - neuro-symbolic AI is one of them
  - need system architectures (e.g. SOFAI, Rossi et al.)

References I

Franz Baader and Bernhard Hollunder.
KRIS: knowledge representation and inference system.
SIGART Bull., 2(3):8–14, 1991.

Alexander Borgida, Ronald J. Brachman, Deborah L. McGuinness, and Lori Alperin Resnick.
CLASSIC: A structural data model for objects.
In James Clifford, Bruce G. Lindsay, and David Maier, editors, *Proceedings of the 1989 ACM SIGMOD International Conference on Management of Data, Portland, Oregon, USA, May 31 - June 2, 1989*, pages 58–67. ACM Press, 1989.

Ronald J. Brachman and James G. Schmolze.
An overview of the KL-ONE knowledge representation system.
Cogn. Sci., 9(2):171–216, 1985.

Scott E. Fahlman.
NETL, a system for representing and using real-world knowledge.
MIT Press, 1979.

Richard Fikes and Nils J. Nilsson.
STRIPS: A new approach to the application of theorem proving to problem solving.
Artif. Intell., 2(3/4):189–208, 1971.

Daniel Kahneman.
Thinking, fast and slow.
Penguin, London, 2012.

Douglas B. Lenat and Edward A. Feigenbaum.
On the thresholds of knowledge.
In John P. McDermott, editor, *Proceedings of the 10th International Joint Conference on Artificial Intelligence. Milan, Italy, August 23-28, 1987*, pages 1173–1182. Morgan Kaufmann, 1987.

References II

Allen Newell and Herbert A. Simon.
Computer science as empirical inquiry: Symbols and search.
Commun. ACM, 19(3):113–126, 1976.

Christof Peltason.
The BACK system - an overview.
SIGART Bull., 2(3):114–119, 1991.

Stuart C. Shapiro.
The sneps semantic network processing system.
In Nicholas V. Findler, editor, *Associative Networks*, pages 179–203. Academic Press, 1979.

Yoav Shoham.
Agent-oriented programming.
Artif. Intell., 60(1):51–92, 1993.

John F. Sowa.
Conceptual graphs for a data base interface.
IBM J. Res. Dev., 20(4):336–357, 1976.

John F. Sowa.
Conceptual Structures: Information Processing in Mind and Machine.
Addison-Wesley, 1984.

Alan M. Turing.
Computing machinery and intelligence.
Mind, LIX(236):433–460, 1950.

References III

Kai von Luck, Bernhard Nebel, Christof Peltason, and Albrecht Schmiedel.
BACK to consistency and incompleteness.
In Herbert Stoyan, editor, *GWAI-85, 9th German Workshop on Artificial Intelligence, Dassel/Solling, Germany, September 23-27, 1985, Proceedings*, volume 118 of *Informatik-Fachberichte*, pages 245–256. Springer, 1985.