CHIIR Tutorials_Agentic AI IR

Information Seeking in the Age of Agentic AI

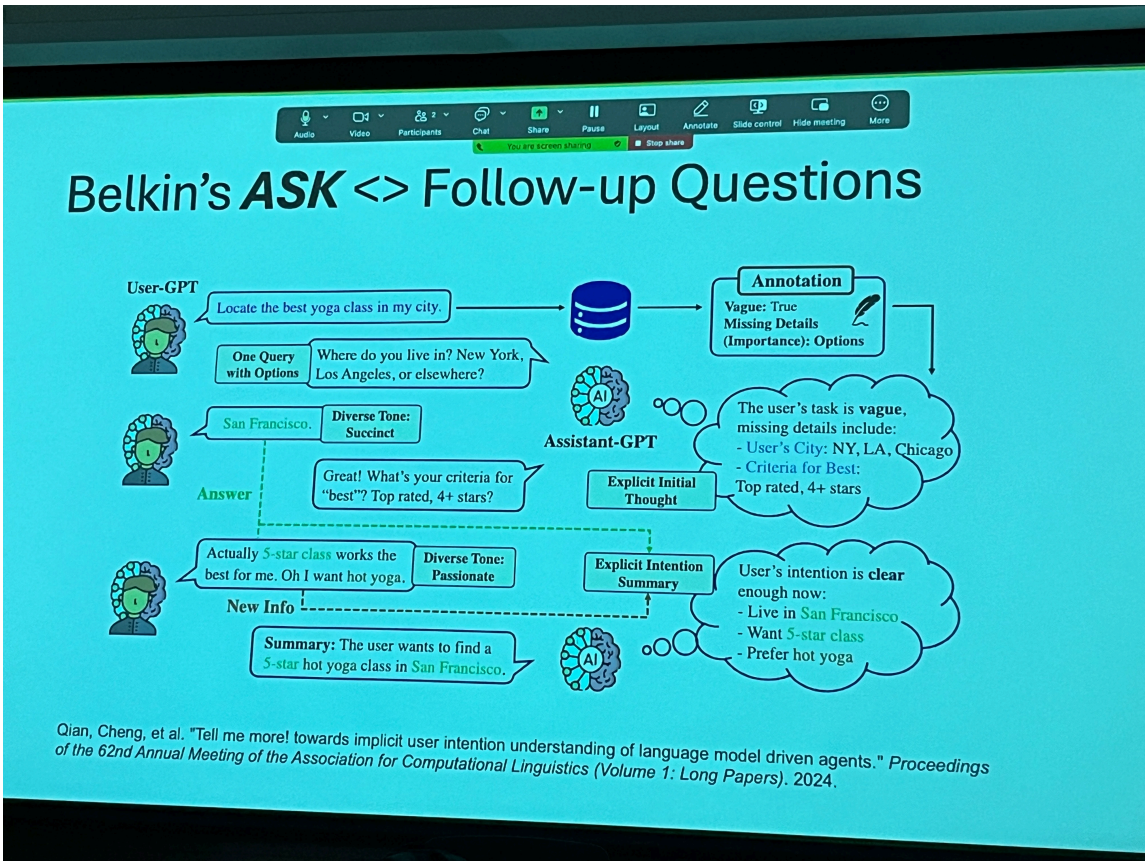
(Preetam Dammu, Tanya Roosta)

information need → information landscape → system design

- what they try to find? what would satisfy their needs?
- what sources? what structures?

Belkin's. anomalous state of knowledge (ASK)

- non-specifiability of information need
- request is incomplete, distorted expression of the underlying need
- different retrieval strategies depending users



Qian, Cheng, et al. "Tell me more! towards implicit user intention understanding of language model driven agents". 2024

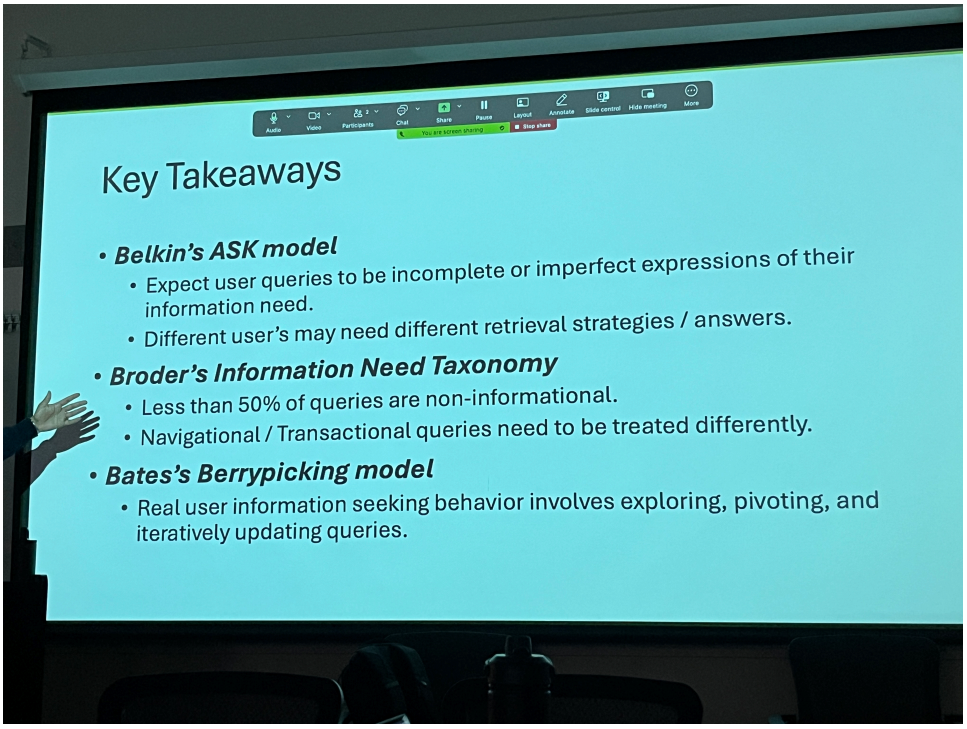
Bate's berrypicking model

- ppl reformulate, pivot, and accumulate 'berries'
- *similar to what agents do. (orchestration process)*

Broder's Information need taxonomy

- search queries are not informational
- three way taxonomy
  - navigational; immediate intent to reach a particular site
  - transactional; not trying to learn but perform web-mediated activity (ex. shopping)
  - informational; acquire some information (abstract, wide funnel..)

→ users do not specify their needs in fleshed outcome



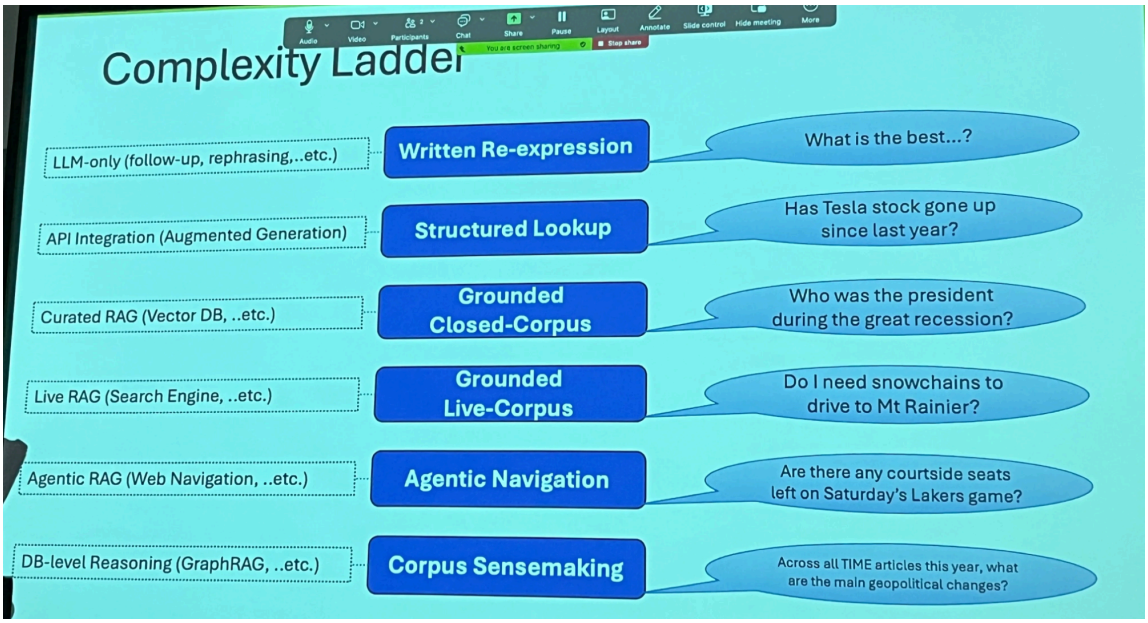
HOW LLMS help?

→ Follow-up questions

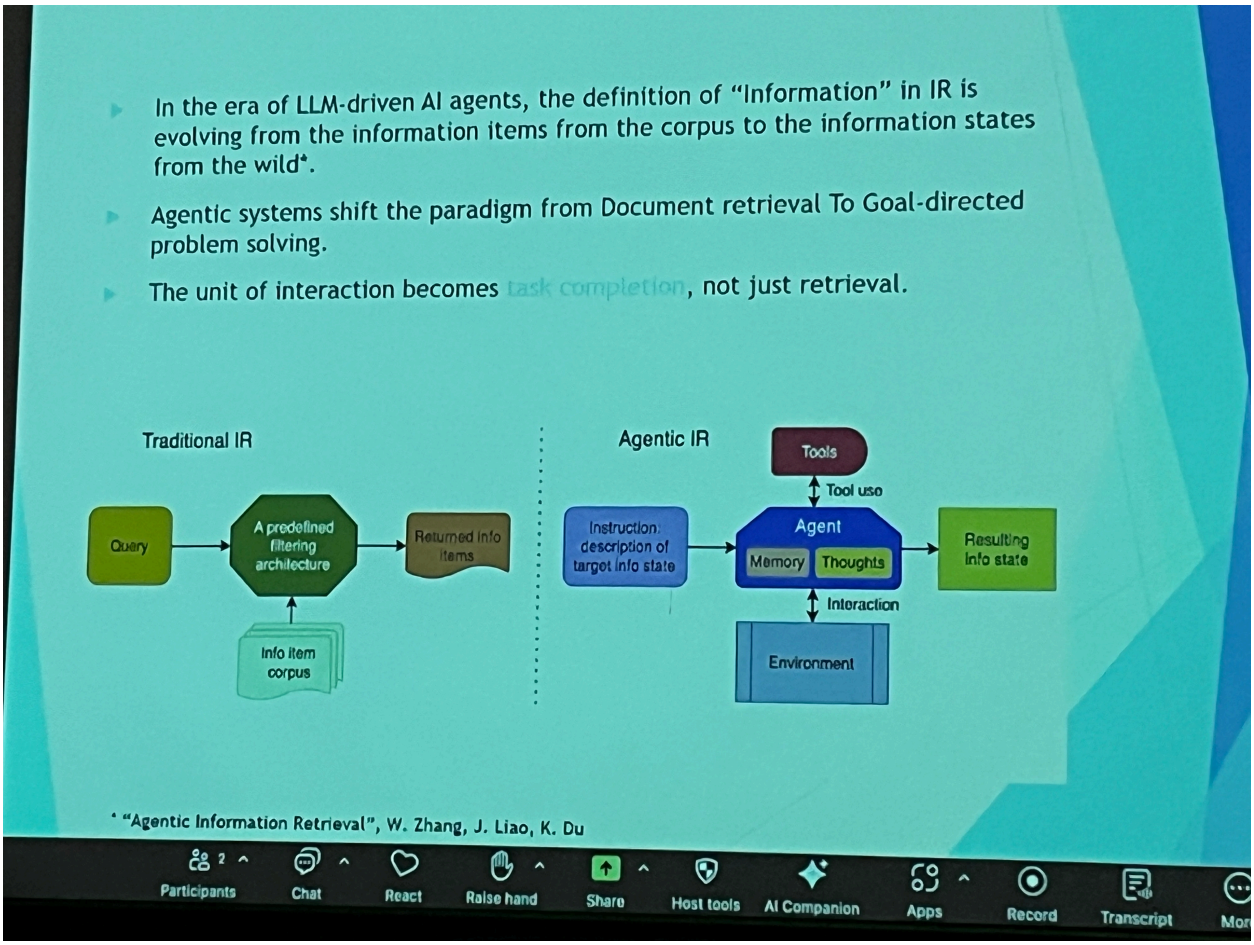
- towards implicit user intention
- agents going through thought steps

→ Agentic Exploration

- **complexity ladder**



Agentic AI for Inf. Retrieval



Agentic Information Retrieval, W. Zhang, J. Liao, K. Du

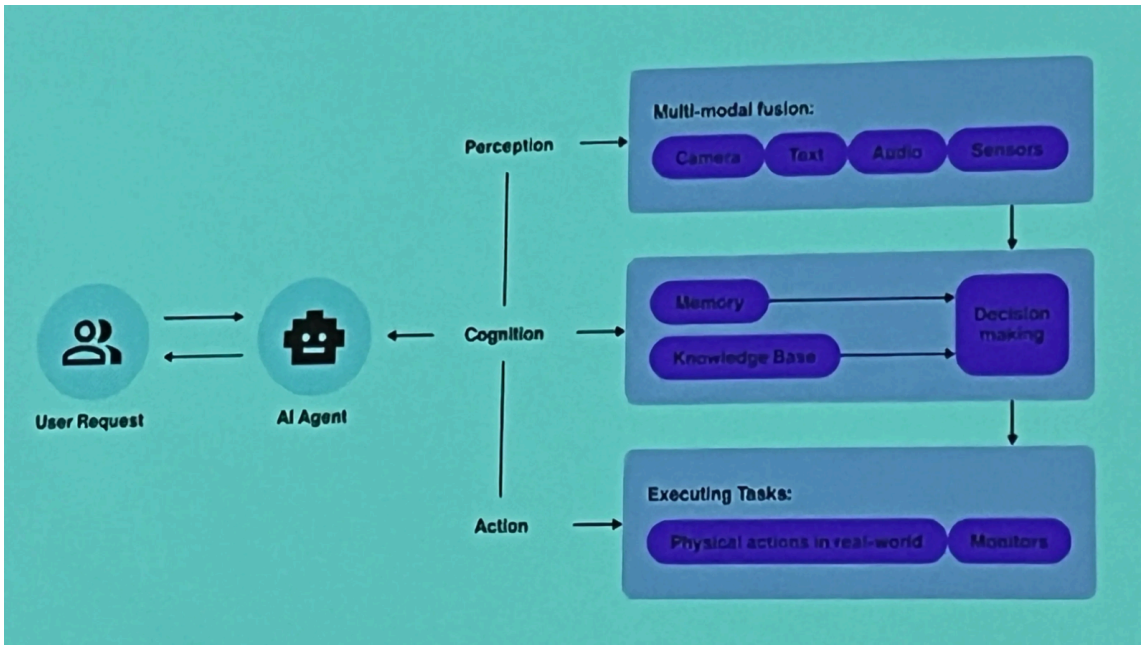
IR traditionally

- query understanding
- document ranking
- relevance estimation
- query > ranked lists > inspecting results > clicking documents > reformulates query until satisfactory answers

⭐ BUT Actually real-world,

- **multi-step** reasoning
    - task decomposition

    - planning and subgoal decomposition
- **comparison** of sources
    - tool choices and orchestration (how to best retrieve information)
    - dataset contamination challenges
        - pretraining + retrieval overlap..
        - multi-step hide leakage..
    - inspectability and auditability
        - faithful vs. intuitive
- **synthesis** of evidence
    - initiate lookups and corroborate data
    - context management
        - memory summarization, retrieval filtering, context pruning
    - memory management
        - balancing between explicit vs. compressed (embedding based)
- from reactive/passive TO GOAL-SEEKING, DYNAMIC
- breaking down objectives and execute actions across multiple steps

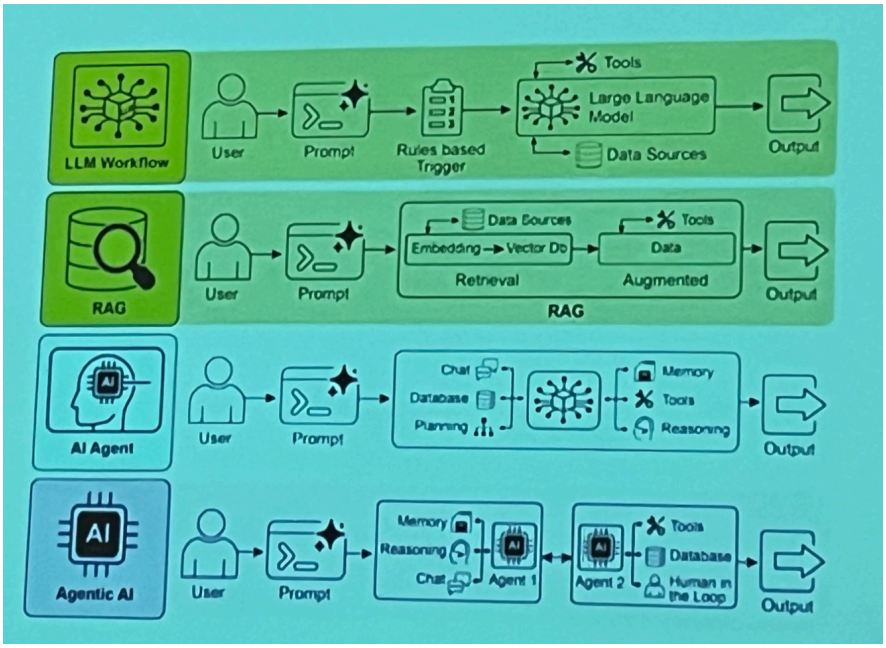


- Agentic systems transform information retrieval by shifting from:
  - queries → goals
  - documents → answers with evidence
  - click logs → interaction traces
- For IR researchers, this introduces new opportunities in:
  - Evaluation
  - User interaction studies
  - System design

Agentic AI

- beyond reactive

- proactive and adaptive (Self-learning)
- perceive, reason, decision making, act autonomously
- contextual awareness → personalization
- temporal drift
    - changes in underlying information over time that make previously correct answers outdated (stored memory becoming stale)
- focus on “accomplishing the goals” ≠ RL (optimal decision making)
- Cognitive Architectures: integration of memory, planning, self-reflection



- reactive agents
    - straightforward execution, respond to current stimuli (no internal memory)
- rational agents
    - using internal models to plan actions
- learning agents
    - improve performance over time through experiences

- BDI (belief- desire-intention) agents
  - simulated based to committed plans
- embodied agents
  - physical, virtual body in simulated world
- LLM based agents
  - LLMs orchestrating actions using APIs/tools
  - alignment, hallucination, evaluations come to challenges

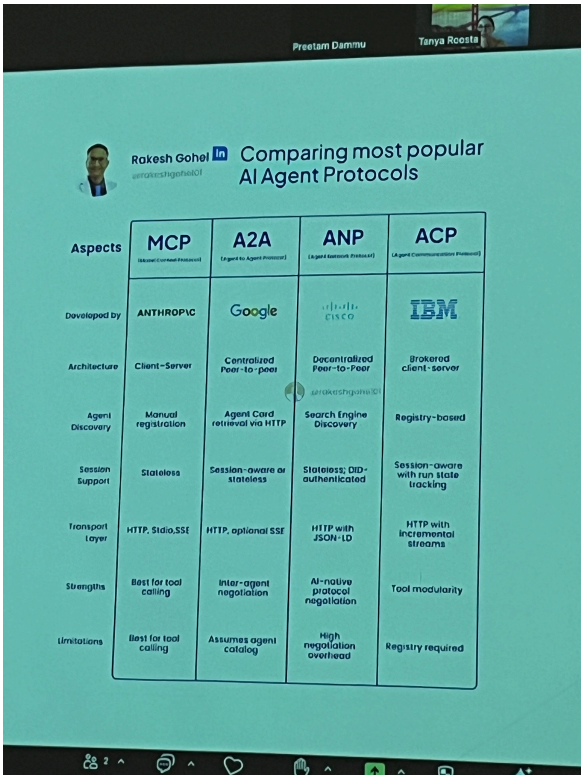
AI agent (working on behalf of user) ≠ Agentic AI (multiple agents interacting)

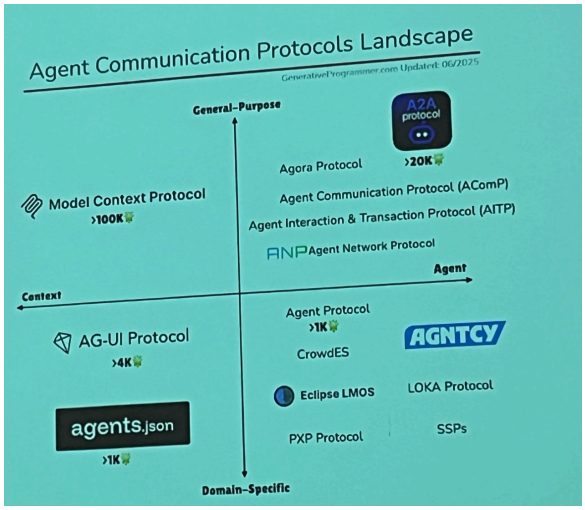
- profiling
  - assigning agent identity, objectives, boundaries, guidelines for interactions with users and systems
- knowledge
  - domain-specific expertise, helping better decision-makings
  - semantic memory, instructions for handling queries
- memory
  - store and retrieve interaction-specific data
  - input from previous, past queries → outcomes → adaptable
  - semantic memory, procedural, ...?
- reasoning and planning
  - sequence of actions to ensure a reliable outcome
  - complex situations, strategic planning
  - chain-of-thoughts, subgoal decomposition
- reflection
  - process feedback from the user and unit tests tools

- actions
    - function calling with APIs, tools to perform tasks

Agent Protocols

- standardized frameworks for structured communication among agents, and between external systems
- interoperability + connected network of intelligence
- MCP, A2A, ANP, ACP protocols





Evaluations of agentic AI

- clear measurable objectives, indicators → quant/qual data →

    - decomposing complex tasks
    - routing pattern; implementing conditional logic

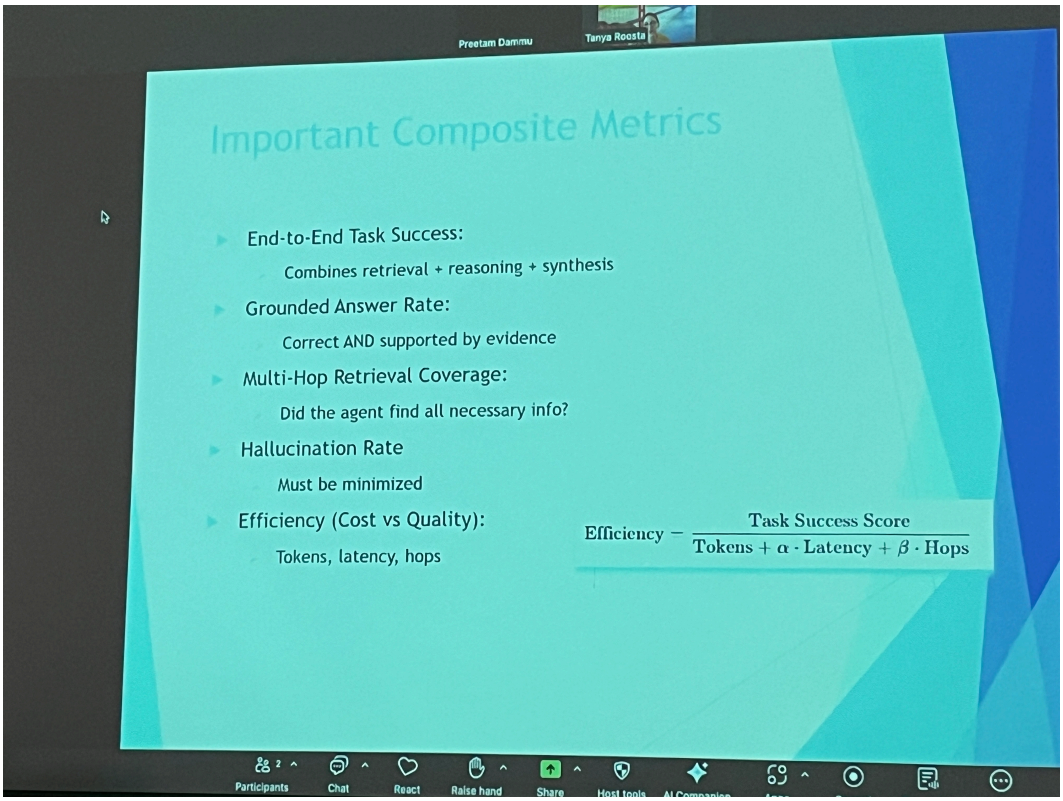
5 dimensions

- autonomy
- goal alignment
- adaptability and learning
- transparency and explainability
- safety and robustness

Metrics

- behavioral benchmarks
- simulation environments
- human in the loop
- agentic metrics
    - initiative index
    - delegation score

NEW METRICS



- correctness
- evidence support
- adequacy
- efficiency
- multistep process
  - task success rate
- faithful metrics
  - how much backed by the sources
- source quality (especially in open-domain)
  - source authority score
  - source freshness
  - viewpoint diversity
- freshness

    - staleness error rate
    - temporal consistency

→ from model weights to entire decision process

→ from predictions to actions with consequences

→ from static evaluation to dynamic trajectories

> not just keyword matching but converting query to decision problem

Further Research Areas...

- memory design for long horizon IR agents
- governance of agentic systems
- user interaction traces
- system design

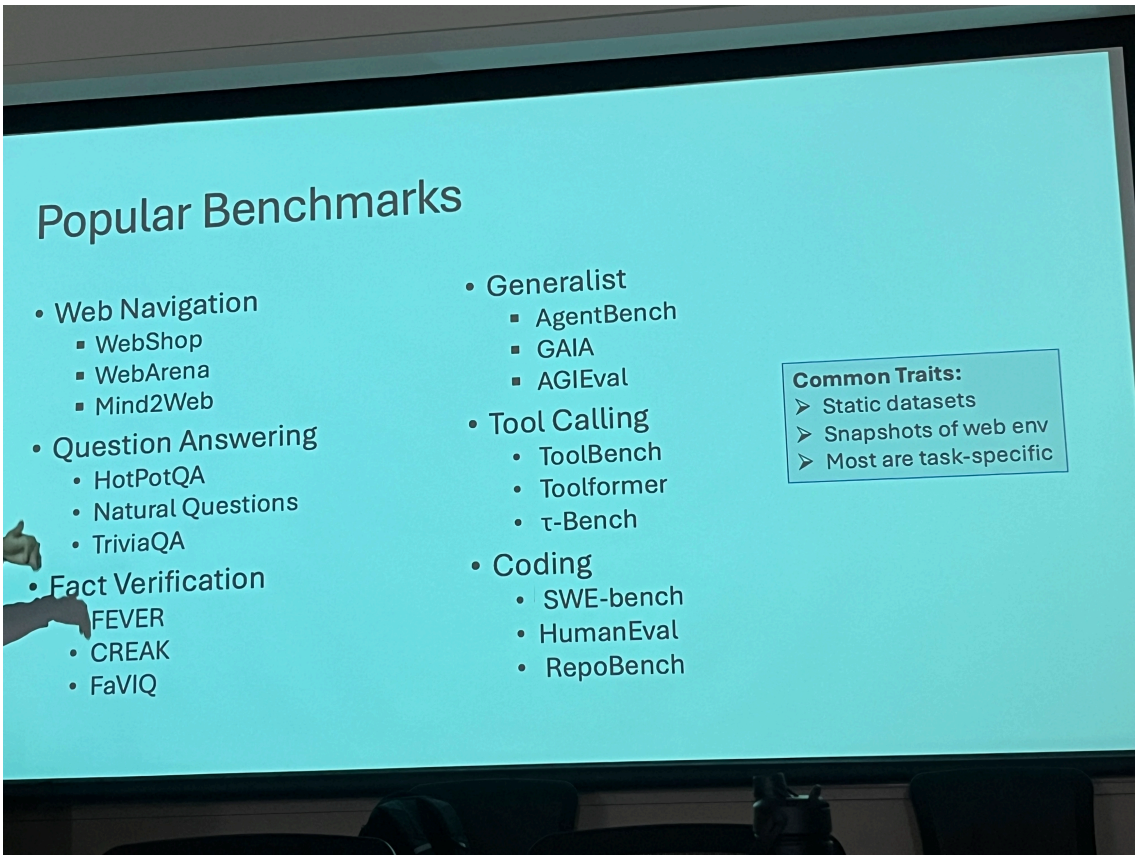
https://isa-tutorial.github.io/isa-tutorial/activity1.html

- live corpus; require simple web pages
- agentic navigation; require multiple steps and buttons to find the information (rely on UI)

RAG = LLM + external info (SQL, knowledge graphs, search engines, etc..)

Evaluation Agentic Information Seeking

- benchmarking; using predefined sets of metrics to evaluate the perfomance
    - web navigation, fact verification, tool calling, coding... → under specific tasks



- traditional ML → deep learning (Transfer learning) → Transformers
  - IR can be contaminated....! with web-access of RAG
- Information Contamination can happen
  - during pre-training
  - or retrieval
- Dynamic Benchmarking
  - adapts test samples at eval time to avoid test contamination
  - models are scored on fresh, unseen data.
  - but still have stable or controllable complexity + consistency across

simulation : behavioral, cognitive fidelity

simulation is not wrong. what are you going to do about it?