Agentic AI Projects

Anupam Datta (Snowflake)
John Mitchell (Stanford)

Outline for today

- Course projects
  - Teams due next week; format for class presentation
- Evolving power of AI
  - Historical context: small model, LLM, Chat, Agents
  - How can we invent the future?
- Project ideas
  - Examples and topic suggestions

Course Projects

team due next week

Project pitches

- Sign up for time slot (Tues/Thurs) by Sunday midnight
  - Place draft slide with your name in slide set by Sunday
- Present your plan in class next week
  - Required for everyone in class
  - Put one slide in a slide set before class
  - Feedback from teaching team and class
  - Cannot all go Thursday
    - 30 teams -> 15 presentations per class
    - 1 min to read your slide; 2-3 min feedback

Reddit personal finance research

- Team members: Allison, John
- Functionality
  - Answer user questions by researching reddit database
  - Agent will look for similar questions and compare reddit user response
  - Agent will summarize reddit responses; discuss with user
- Evaluation metrics
  - Relevance of selected reddit samples
  - Ability to rank reddit comments based on discussion context
  - Accuracy and completeness of actual summary, relative to reddit samples used to generate response

Starting point for planning your project

Agentic systems

- Plan ways to proceed
- Select tools
- Call tools
- Combine results
- Synthesize information
- Critique before responding

Sample applications

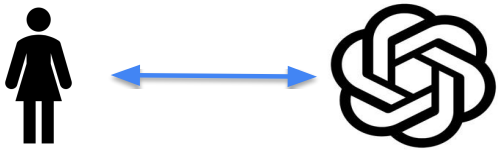
- Research a topic and respond
  - Finance
  - Health
  - Science, History, …
- Plan a trip, including
  - Locations to visit
  - Hotels and restaurants
  - Within time and budget
- Perform a complex task
  - Decide what actions needed
  - Carry out complex steps
  - Combine/synthesize results
- …

Evolving power of AI

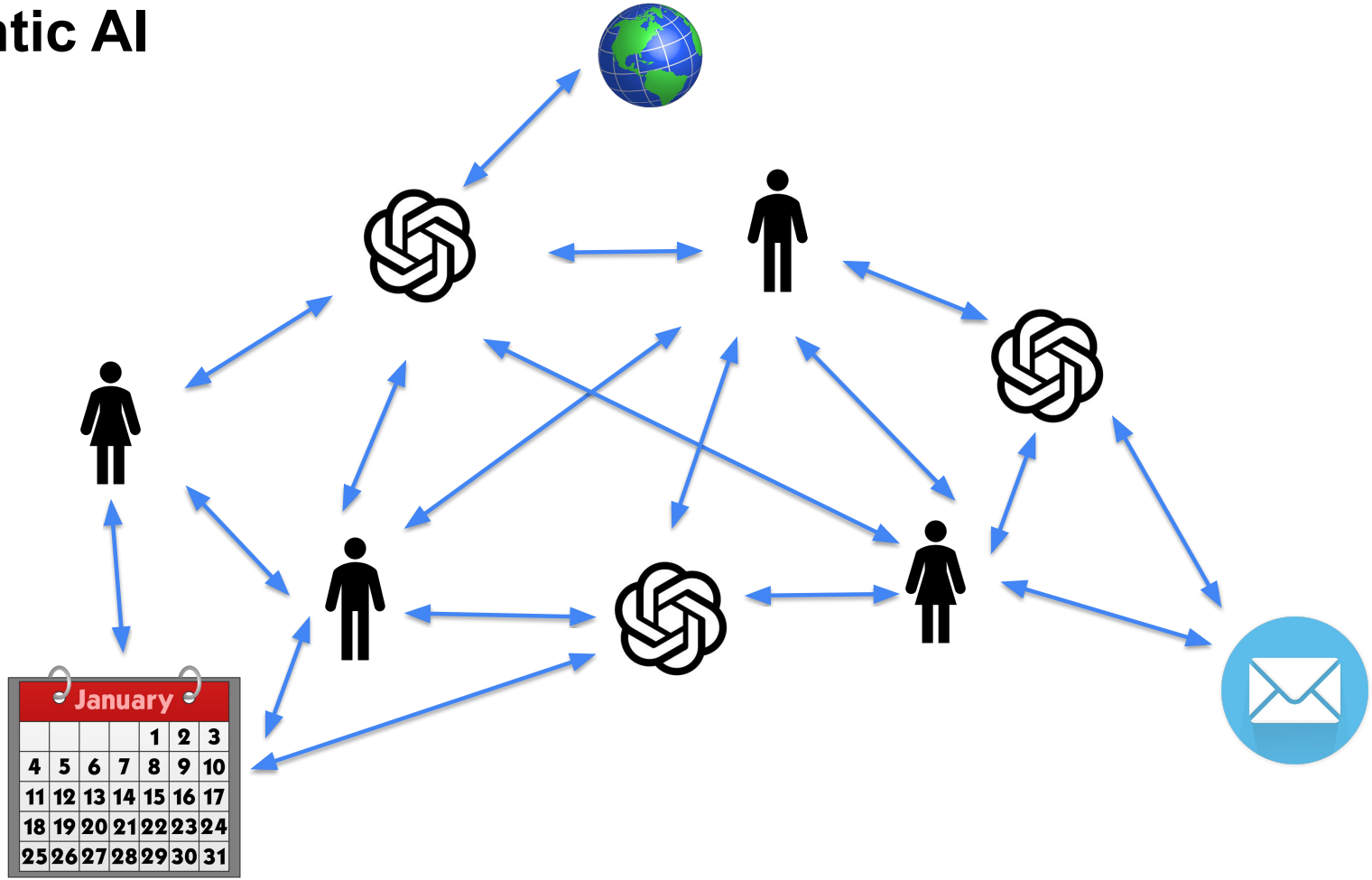
Big Picture: How can we use AI in 2026?

- Pre-LLM
  - Collect data, train a model, use AI for classification: loan application, parole, …
- ChatGPT
  - Makes stuff up, sounds intelligent, inspired everyone to think AI will solve everything
  - Build useful initial applications based on prompting
- Retrieval-Augmented Generation (RAG)
  - Collect useful data, answer questions using LLM to interact with user,
- Agentic systems
  - Structure system of agents based on traditional human workflow
  - Good performance in specific task areas: software dev, …
- Future?
  - How do we leverage human expertise to build AI agents that make productive contributions to organized group work?

Chatbot



Agentic AI


January
1 2 3
4 5 6 7 8 9 10
11 12 13 14 15 16 17
18 19 20 21 22 23 24
25 26 27 28 29 30 31

Example: Code development using LLMs

- Human + LLM assistant
  - Github Copilot
- LLM alone
  - Ask LLM to write the program
- Team of AI agents
  - Assign different roles to each agent
  - Structure the code development process
- Team of humans and AI agents
  - Allow humans to assume specific roles in collaboration with AI agents in other roles

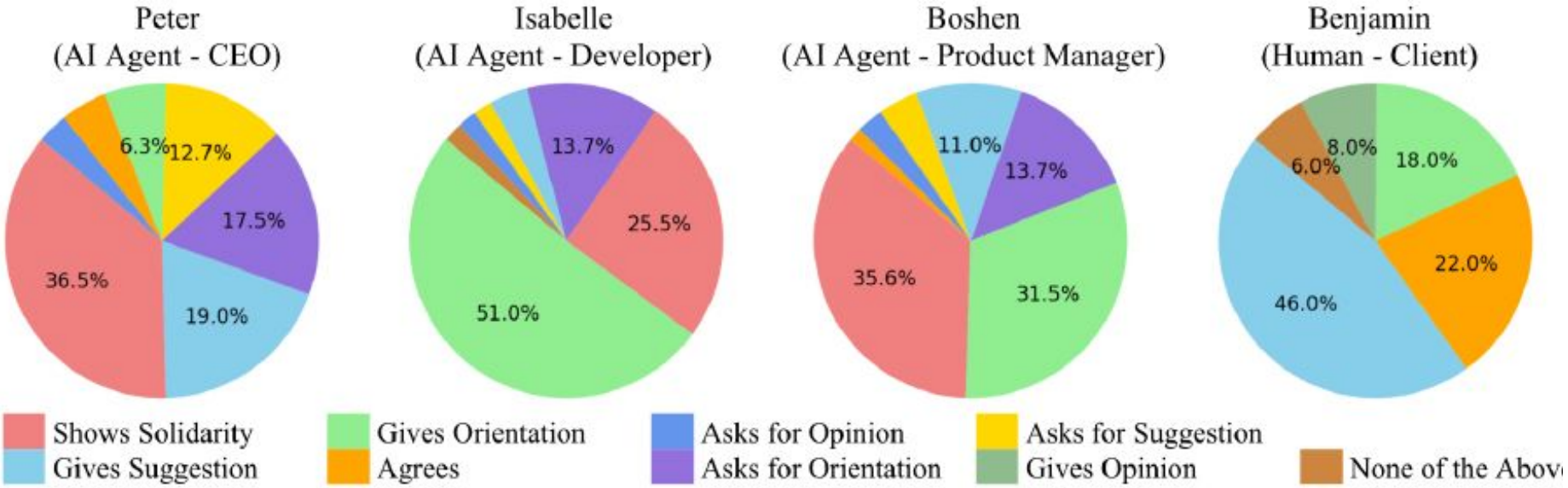
ChatCollab

multi-agent human-AI collaboration

- Agents may be human or AI
  - Assign each agent a role
  - Allow workflow to proceed accordingly
- System could be used for many purposes
  - Assign roles that make sense for specific setting
  - Prompt AI agents to help eah fill their role
- Experiments so far consider software dev
  - Compare to multi-AI-agent systems w/o humans
  - Transcript analysis to evaluate collaboration

	ChatGPT	SuperAGI	MetaGPT	ChatDev	ChatCollab
INTERFACE					
Interactive with user in each step	✗	✗	✗	✗	✓
FUNCTIONALITY					
1. The code compiles without errors	✓	✓	✓	✓	✓
2. Uses 'X' and 'O' for the two players	✓	✓	✓	✓	✓
3. Creates a 3x3 grid	✓	✓	✓	✓	✓
4. Guides the players through the game	✓	✓	✓	✓	✓
5. Starts the game by displaying an empty board	✓	✓	✓	✓	✓
6. Starts the game by assigning 'X' to the first player and 'O' to the second player	✓	✓	✓	✓	✓
7. Prompts the players to input their moves by specifying the row and column	✓	✓	✓	✓	✓
8. Handles non-integer input	✗	✗	✗	✓	✓
9. Ensures that the user input is not out-of-range	✓	✓	✓	✓	✓
10. Ensures that the user input is not in an already occupied cell	✓	✓	✓	✓	✓
11. Correct placement of X's and O's according to user input coordinates	✓	✓	✓	✓	✓
12. Displays the updated board after each move	✓	✓	✓	✓	✓
13 .Displays the final board after the game ends	✓	✓	✓	✓	✓
14 .Detects the winner	✓	✓	✓	✓	✓
15. Announces the result of game as soon as a player wins	✓	✓	✓	✓	✓
16. Announces the result of game if it ends in a tie	✓	✓	✓	✓	✓
17. After the game concludes, asks for new game and if so restarts the game	✗	✗	✓	✓	✓
CODE QUALITY					
1. Decomposition	✗	✓	✓	✓	✓
2. Source Code Documentation (general comments, inline comments, etc.)	✗	✗	✗	✓	✓
3. Supporting material (user instructions, summary, notes, etc.)	✗	✓	✓	✓	✓

Comparison of agent behavior



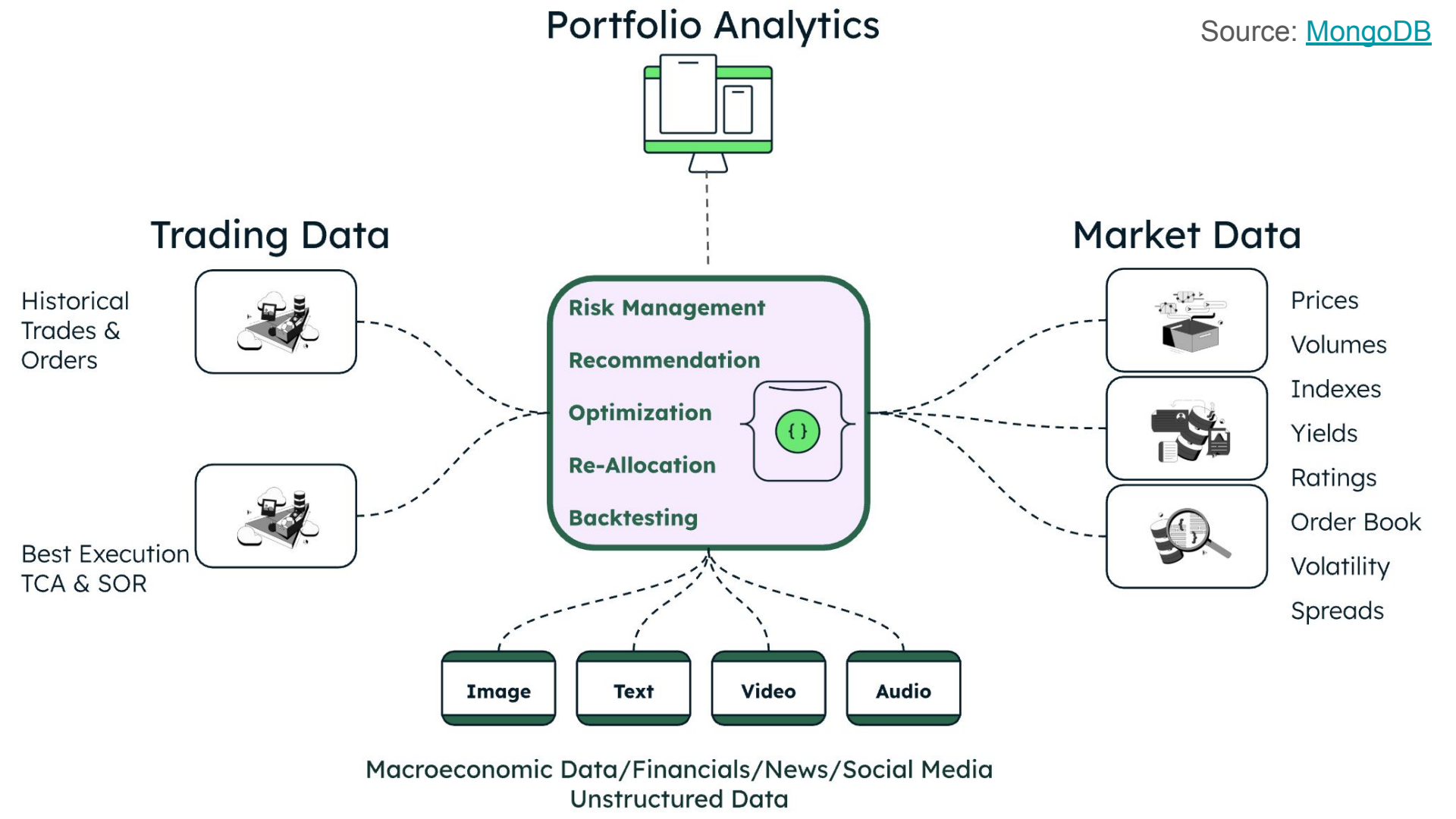
ChatCollab: Exploring Collaboration Between Humans and AI Agents in Software Teams

Question: How can we build ***better*** agentic systems?

- Straightforward approach
  - Identify set of agents needed to accomplish workflow
  - Define a planning agent and an orchestrator
  - Use LLM-judge and other techniques for evaluation metrics
  - Iterate to produce a reasonable, reliable agent for meaningful tasks
- But is there something better?
  - How do we make agents *effective* at their tasks?
    - Prompt them with instructions, based on instructions to humans?
    - Add information we give humans to train them for roles?

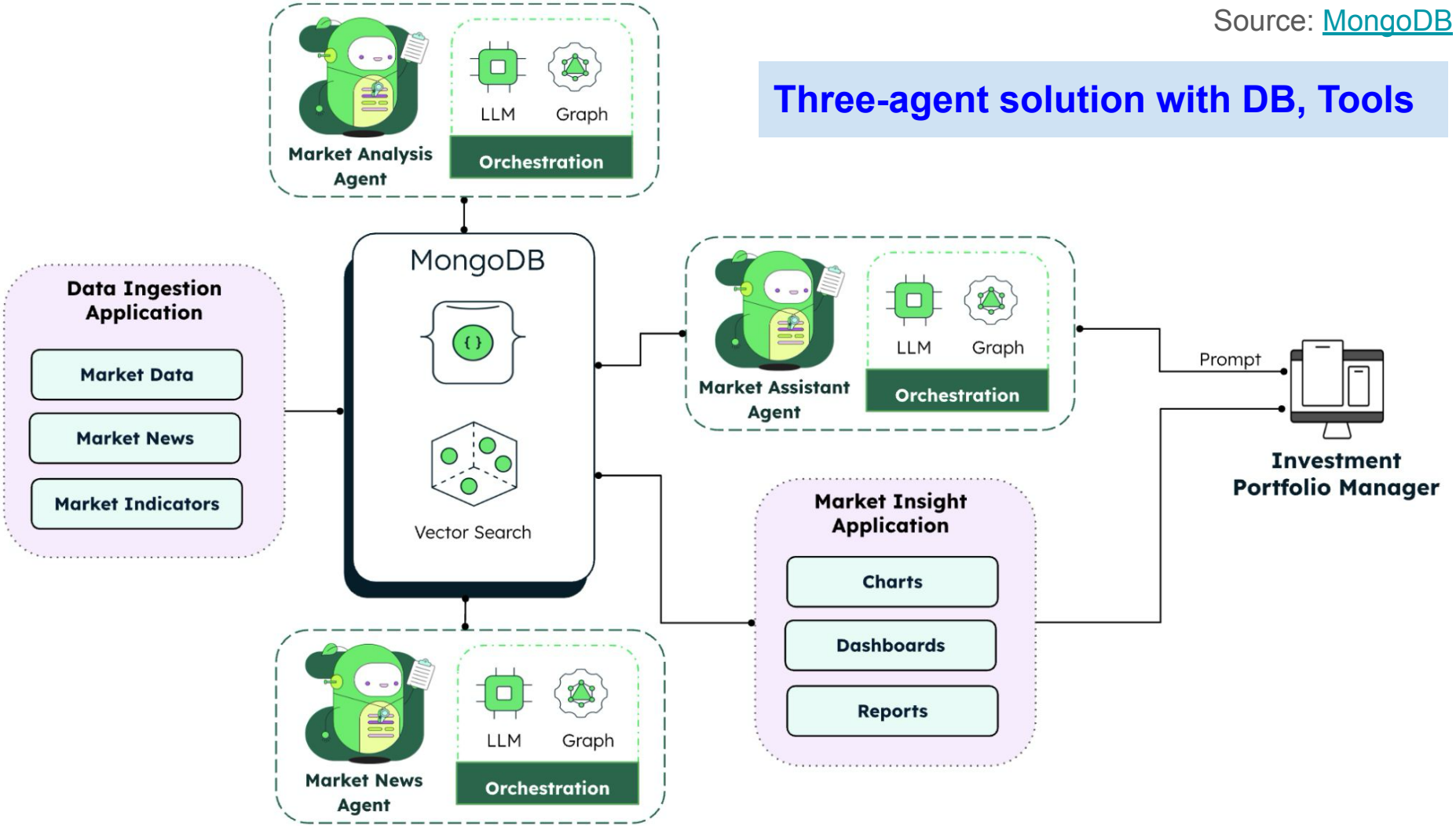
      - **What else could lead to better AI systems?**

Project ideas


Portfolio Analytics
Source: MongoDB
Trading Data
Historical Trades & Orders
Best Execution TCA & SOR
Risk Management
Recommendation
Optimization
Re-Allocation
Backtesting
{ }
Market Data
Prices
Volumes
Indexes
Yields
Ratings
Order Book
Volatility
Spreads
Image
Text
Video
Audio
Macroeconomic Data/Financials/News/Social Media
Unstructured Data

Source: MongoDB

Three-agent solution with DB, Tools



Insurance: *more interesting than you think*

- I bicycle to the coast every weekend
  - I usually go on Saturday, but skip it if I am traveling
  - I have not had a flat tire for 2-3 years, maybe more
- How much would I have to pay you to pick me up if I get a flat tire and cannot fix it easily?
  - $5 a week, assuming a one-year agreement
  - $10/week
  - $25/week
  - $50/week

THE PREMIER MANAGER OF REINSURANCE RISK

Ariel Re offers a range of innovative insurance and reinsurance solutions and services through our offices in Bermuda, London, and Hong Kong, meeting the business needs of a diverse client base. Ariel Re operates through Syndicate 1910 at Lloyd's of London and we also offer access to Lloyd's Europe.

We are a major writer of property catastrophe reinsurance, covering large insurance companies around the world against major losses from natural catastrophes such as hurricanes and earthquakes.

Ariel Re

Island of Bermuda

Two project suggestions from ArielRe

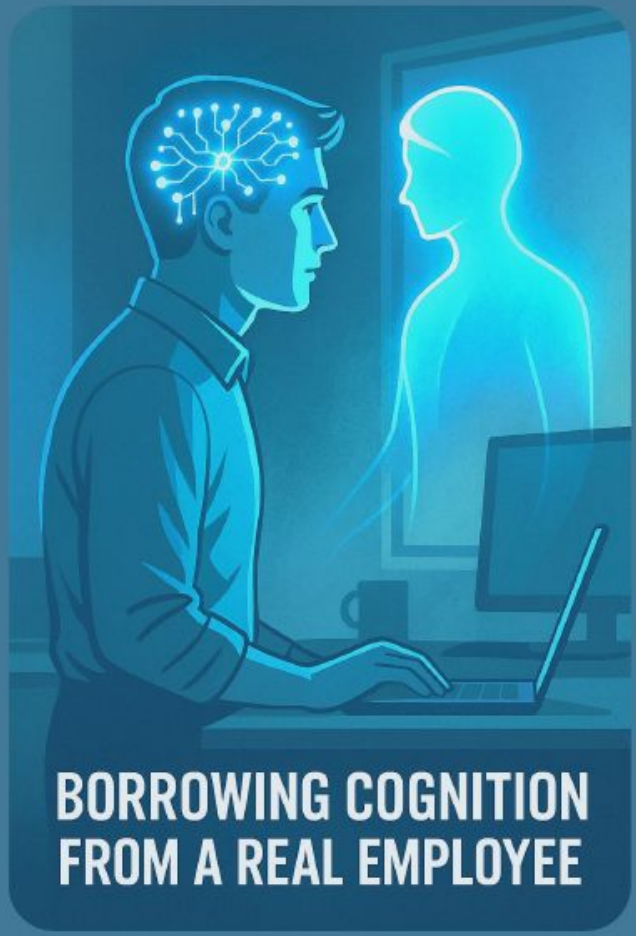
- Agentic AI for Cedant Historical Claim Data Extraction
  - Read reinsurance submission packs
  - Extract historical loss data
  - Link losses to Property Claim Services /internal historical events
  - Use LLM to provide a summary of losses to underwriters
- AI Agent for Reinsurance Contract Terms Validation
  - Validate reinsurance contract terms against data kept in internal systems
  - Extract fields required for policy administration from unstructured documents
  - Compare fields to corresponding internal data in SQL/JSON format
  - Highlight matches, discrepancies, ambiguous data via LLM output

Send email to Jason or me to get 2-page descriptions

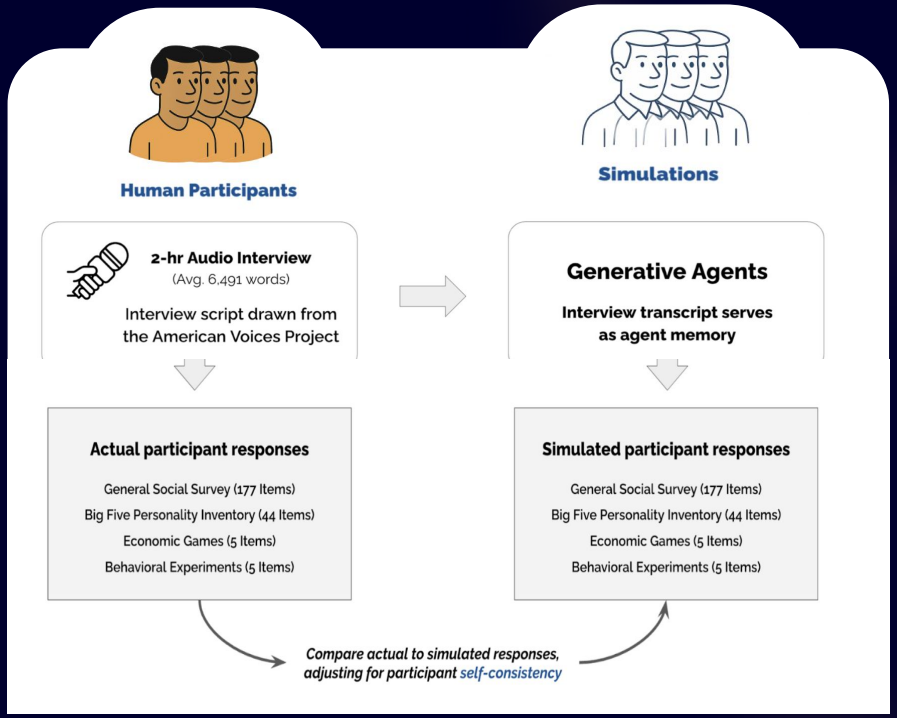
Digital Double Creation through
Dynamic Cognitive Replication

while

Equipping each employee with
the trusted Confidant / Co-Pilot
in a process...



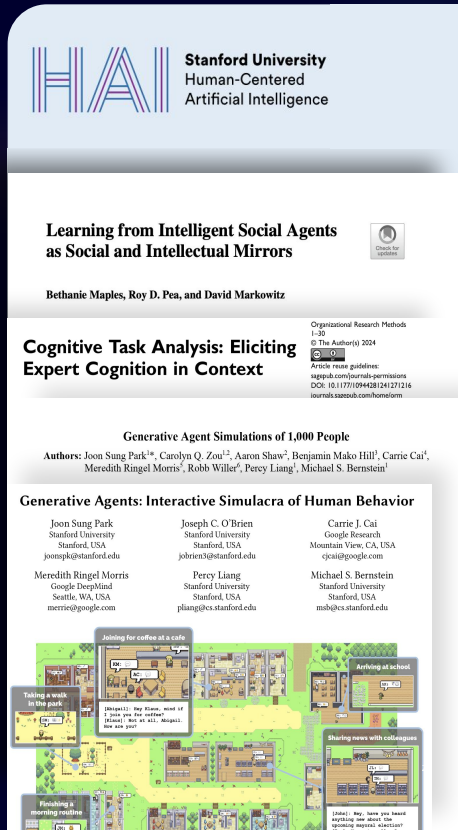
Gen Agents—Cognitive Replicants for Insurance Industry



REPLICATION PROTOCOL

CALIBRATION
COGNITIVE TASK ANALYSIS

VERIFICATION



Park, J. S., Zou, C. Q., Shaw, A., Hill, B. M., Cai, C., Morris, M. R., Willer, R., Liang, P., & Bernstein, M. S. (2024, November 15). *Generative Agent Simulations of 1,000 People.*

Project idea	Your name(s)
sport betting, national security, public safety	Kyle Simmons
Step-wise pub financial research to create ultimate DCF model	Tony Shi
Virtual Classroom with Teacher-Student Agents	David Lyu (dlyu@stanford.edu)
Multi-agent Performance benchmarks versus Single-agent with restraints	Dat Tran (dattran@stanford.edu)
Meal planner based on existing ingredients	Anna Fisher Lopez
video games … maybe	
Agent that can help answer questions about homework or course content based on lecture	Akhil vyas
3D Design Agent.. Video games	Jane Yang (yjane@stanford.edu)
Nutrition - personalized meal planner agent that takes into account desires/needs (dietary restrictions, weight gain/loss, athletic needs, timing constraints for prep, seasonality of food, prices, etc.) and creates meal plan	Michelle Campeau (mcampeau@stanford.edu)
Education project: agent to distill content into digested parts, using science-proven techniques to maximize learning metrics	Julián Rodríguez Cárdenas Arjun Jain

Starting point for planning your project

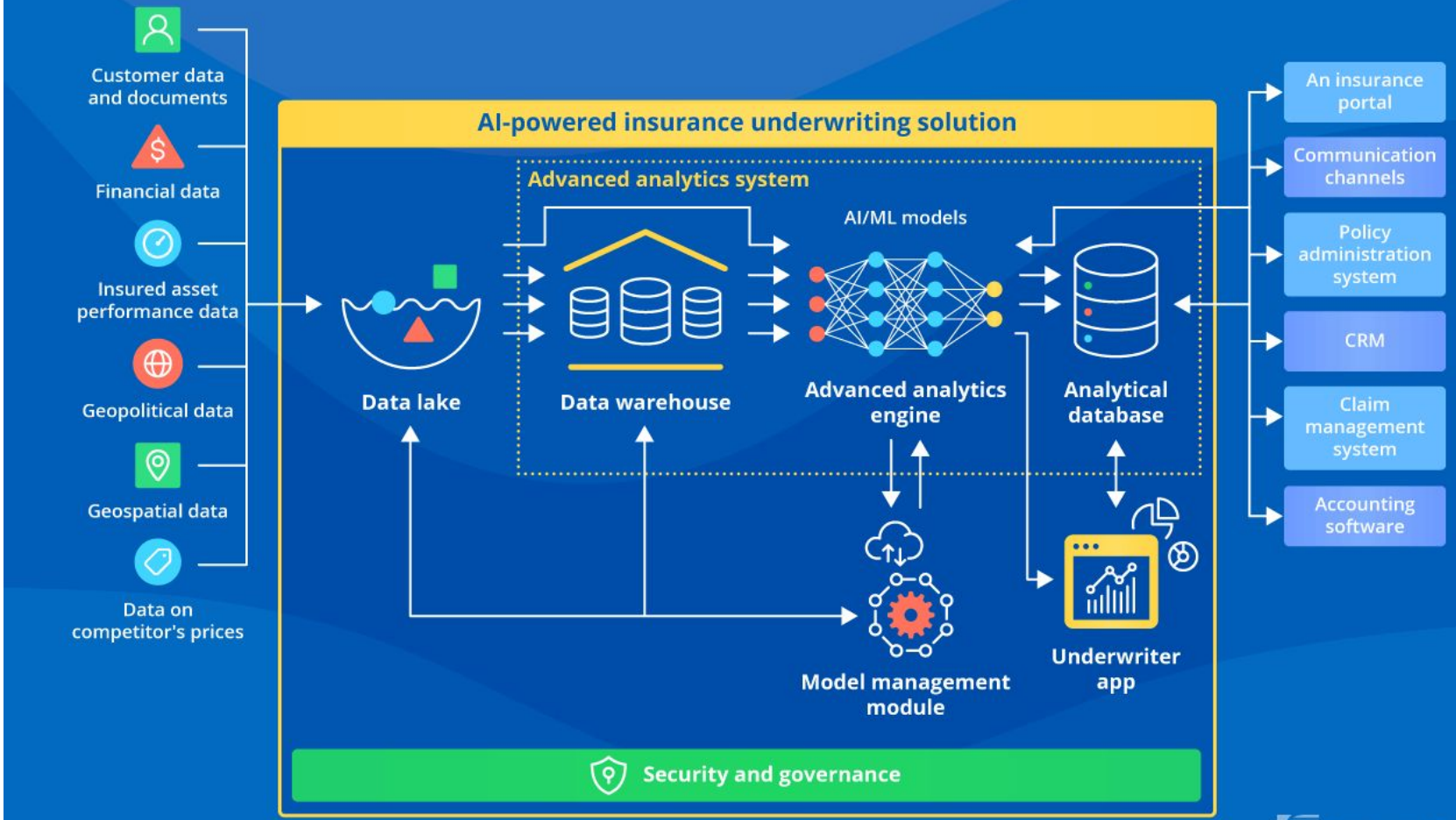
Agentic systems

- Plan ways to proceed
- Select tools
- Call tools
- Combine results
- Synthesize information
- Critique before responding

Sample applications

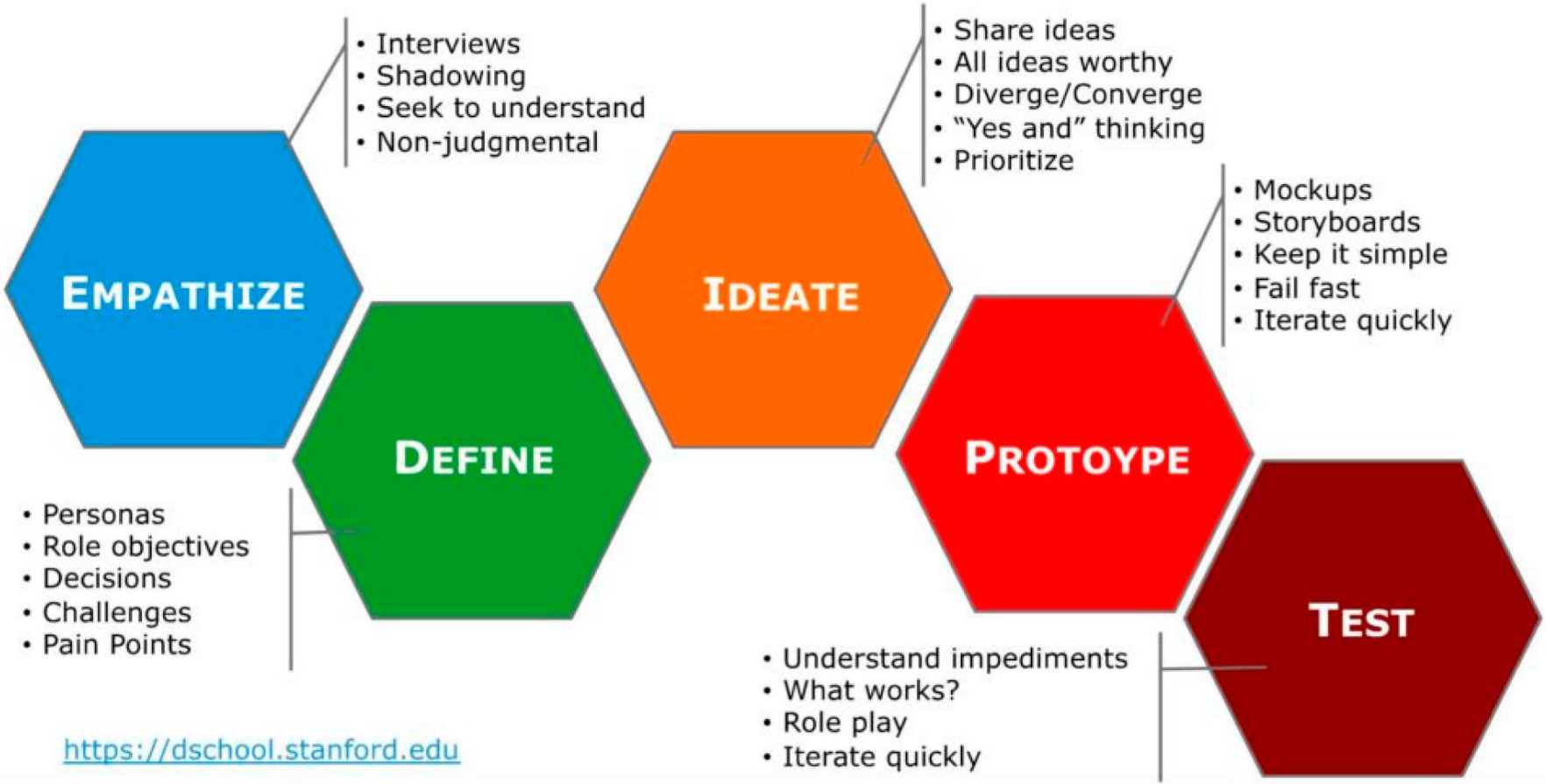
- Research a topic and respond
  - Finance
  - Health
  - Science, History, …
- Plan a trip, including
  - Locations to visit
  - Hotels and restaurants
  - Within time and budget
- Perform a complex task
  - Decide what actions needed
  - Carry out complex steps
  - Combine/synthesize results
- …

Source:ScienceSoft



ScienceSoft

Stanford d.school Design Thinking Process



https://dschool.stanford.edu