DataHop Boston

Open for Innovation
KNIME

Convene
PW: stayconnected

9:30 - 10:15 AM	**Check-in / Morning Tea**
10:15 - 12:15 PM	**KNIME Workshop 1** Enrich Data Analytics with GenAI **KNIME Workshop 2** Deploy a GenAI enriched workflow on KNIME Business Hub
12:15 - 1:30 PM	**Lunch & Networking**
1:30 - 2:00 PM	 **Welcome. Data Science & GenAI** Michael Berthold | KNIME
2:00 - 2:30 PM	 **Streamlining the design of DNA-encoded libraries at GSK using KNIME** Brittany Smith | GSK
2:30 - 3:00 PM	 **Implementation of automated GxP data pipelines using KNIME** Shiva Katepalli | Sumitomo Pharma America
3:00 - 3:20 PM	**Afternoon tea**
3:20 - 3:50 PM	 **How KNIME maximizes scalability and business impact in People Analytics** Christopher Cerasoli | Boehringer Ingelheim
3:50 - 4:20 PM	 **LLMs' RAGs and KNIME for biopharmaceutical applications** Kenneth Longo | Wave Life Sciences
4:20 - 5:00 PM	**Panel: KNIME usage in research** **Moderator:** Sasha Rezvina (KNIME) **Panelists:** Cheng Fang (Blueprint Medicines) Wendy Guan (Harvard University) Giovanni Cianchetta (Recludix Pharma) Michael Berthold (KNIME)
5:00 PM	 **Closing** Michael Berthold | KNIME
5:00 - 6:30 PM	**Networking reception**

KNIME Booth + Worklow Doctor are available during all breaks

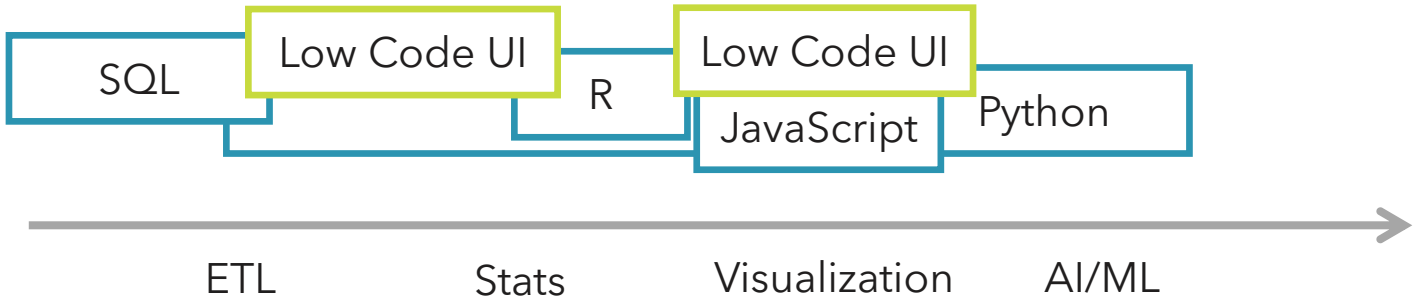
The KNIME stance 2022 and before:

Workflows are the most intuitive way to make sense of data.

Data workers don’t need to understand **how** things are done,
they need to understand **what** is done.

(Low) Code for Data Science: a SW Engineering Practice

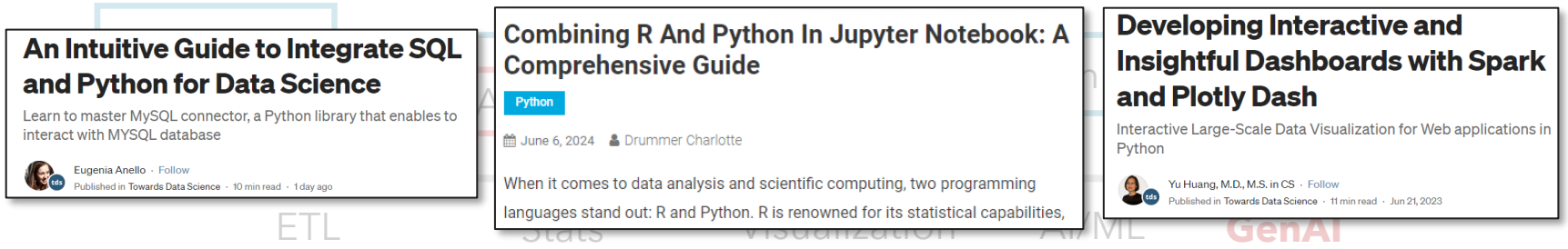
No-code on top of code-based programming languages keeps the complexity



(Low) Code for Data Science: a SW Engineering Practice

The technology your data science team needs to manage:

SQL | R | Python | JavaScript | … | GenAI (APIs)



- Low code UI creates code underneath the hood
- More complex functionality requires adding/editing code
- Complex management & governance of code and dependencies
- New: GenAI adds uncontrolled access to inhouse/cloud resources

Workflows for Data Science: One Consistent Environment

The technology your data science team works with:

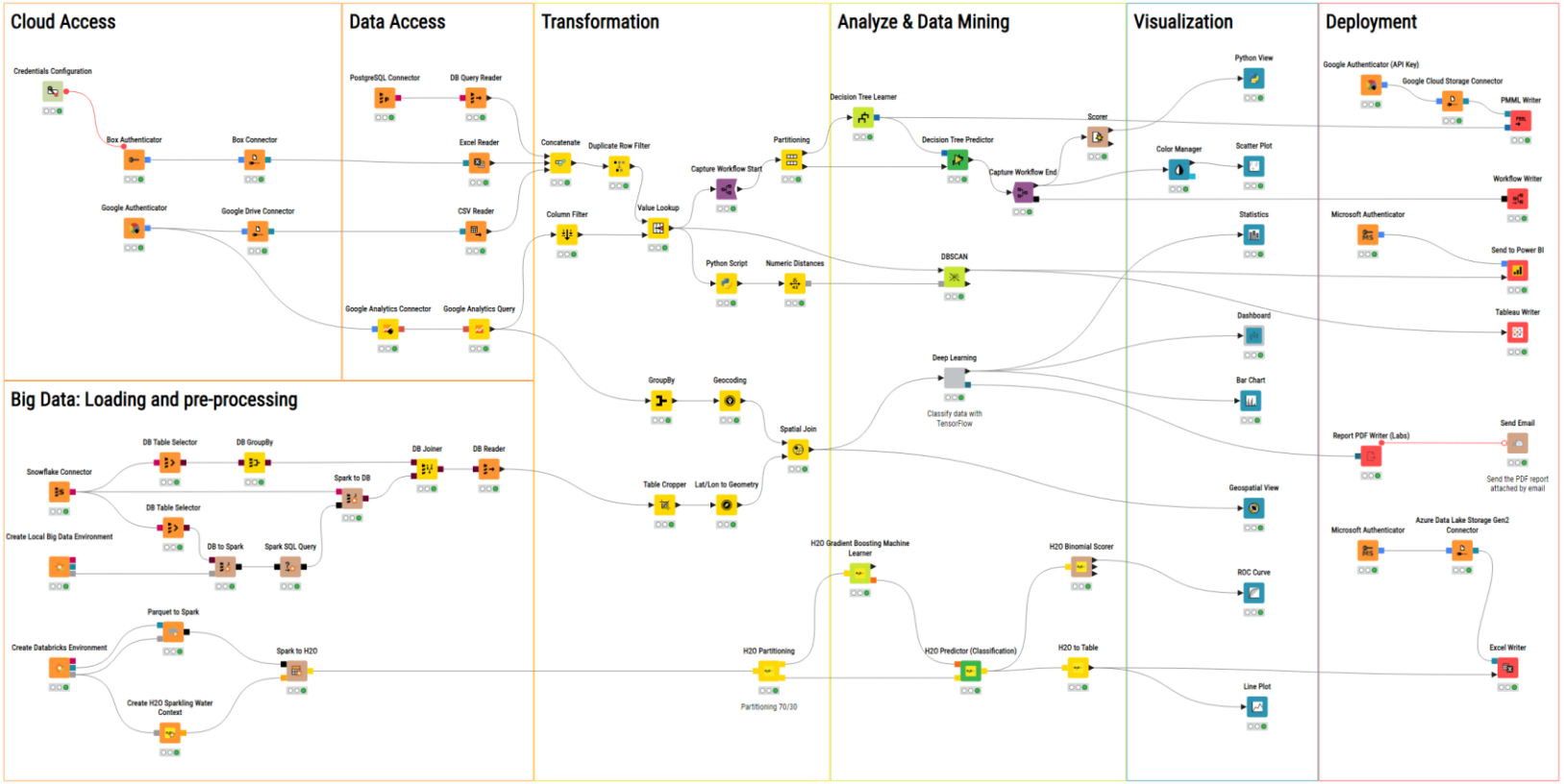
Visual Workflows

- Visual workflows **are** the programming language
- abstracts implementations in various technologies
- allows to embed code snippets
- Governed access to GenAI

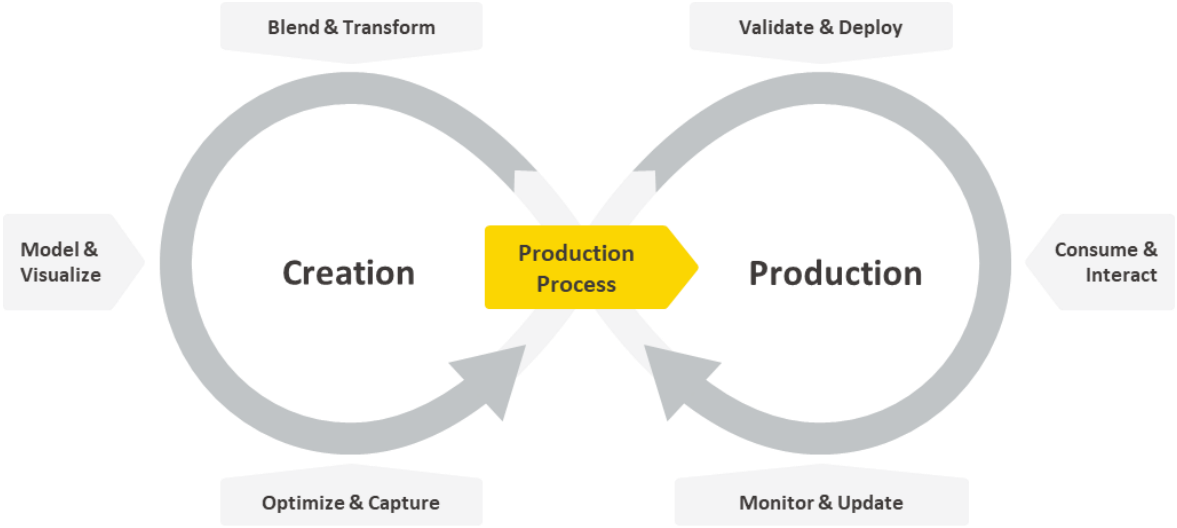
Quick Reminder:
KNIME Workflows & Software

Visual Workflows for Data Science & AI

Provide techniques & capabilities to all types of users



KNIME Supports Every Step of the Data Science Lifecycle



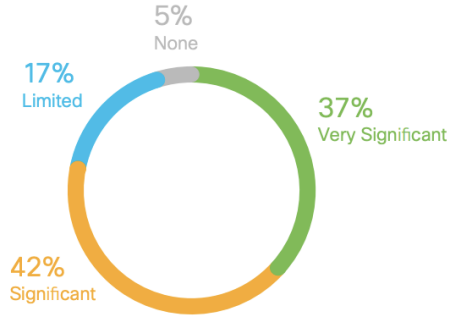
KNIME Analytics Platform (open source)			
Low Code Data Science			
Data Access & Transformation	Analytics & GenAI	Visualization & Reporting	Community Extensions

KNIME Community Hub	KNIME Business Hub			
Team Collaboration	Governance and Deployment at Scale			
Shared Repository	Data Apps & API Services	Automation & Orchestration	Continuous Delivery of Data Science	AI Governance

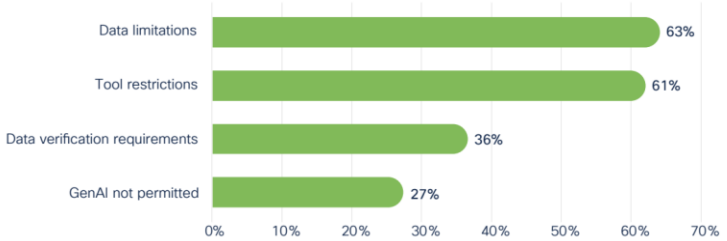
Enter: GenAI…

GenAI: Opportunities and Risks

79% perceive significant value from GenAI adoption[1]



27% of organization have an outright ban on GenAI usage[1]



- \+ Massive impact on “mundane” tasks
- \+ Broad userbase
- \+ Applicability everywhere

Positive impact is blocked…

- ⊖ Questionable quality
- ⊖ Uncontrolled consumption
- ⊖ Governance requirements

…by fear.

source: Cisco 2024 Data Privacy Study

Flexibility or Reliability?

- Typical AI applications are flexible
  - Hallucinations, (fake) creativity, … are ok
  - Works fine for images, starting points, summaries (sometimes), …
- Other AI applications can’t be trusted
  - Hallucinations and fake information are **not ok**
  - careful proofreading required: code, legal docs, …

Flexibility or Reliability?

Working with data has both facets:

- Flexibility is sometimes ok
  - Exploratory data analysis to trigger insights, form new hypotheses, …
  - Nobody cares how those results were generated (code, guessing, wild math…)
- But more often, reliability is required:
  - Accurate data background for reporting, forecasting, …
  - E.g. financial data aggregation for tax reporting & auditing

Workflows explain, allow to validate, and provide the basis for collaboration on how graphs, numerical summaries, detailed forecasts … are created

GenAI Reliability has More to It…

In terms of generated content:

- Correctness
- Transparency (review, document, validate, audit, …)

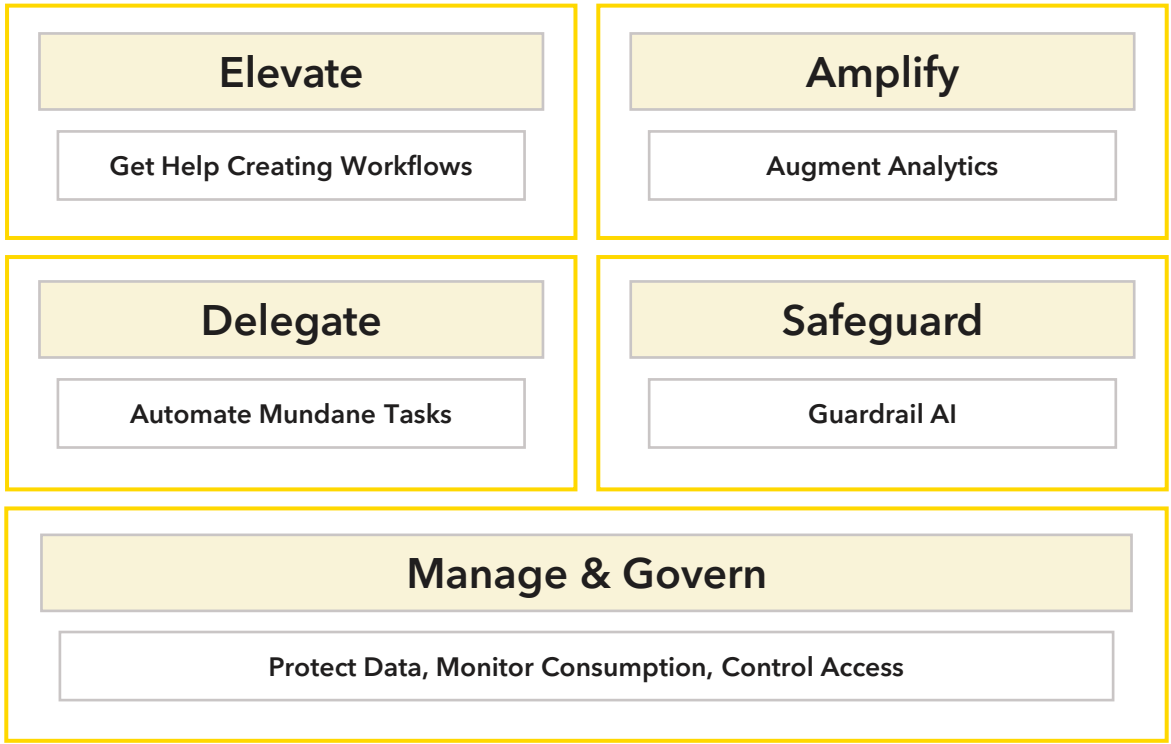
In terms of organization-wide use:

- Governance of models
- Protection of data
- Control of consumption

The KNIME stance in the age of AI:

Workflows are the most intuitive and reliable way to work with data and AI.

Trust AI with Your Data



KNIME and AI

Elevate

Get Help Creating Workflows

Amplify

Augment Analytics

Delegate

Automate Mundane Tasks

Safeguard

Guardrail AI

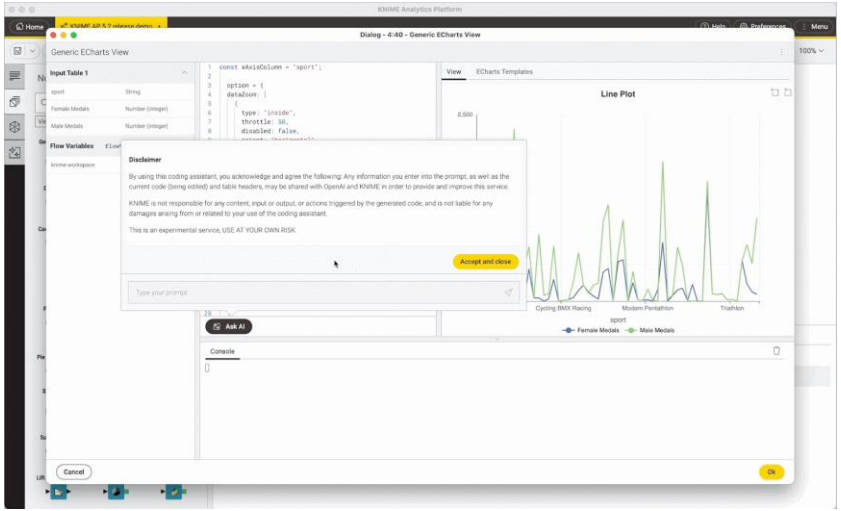
Manage & Govern

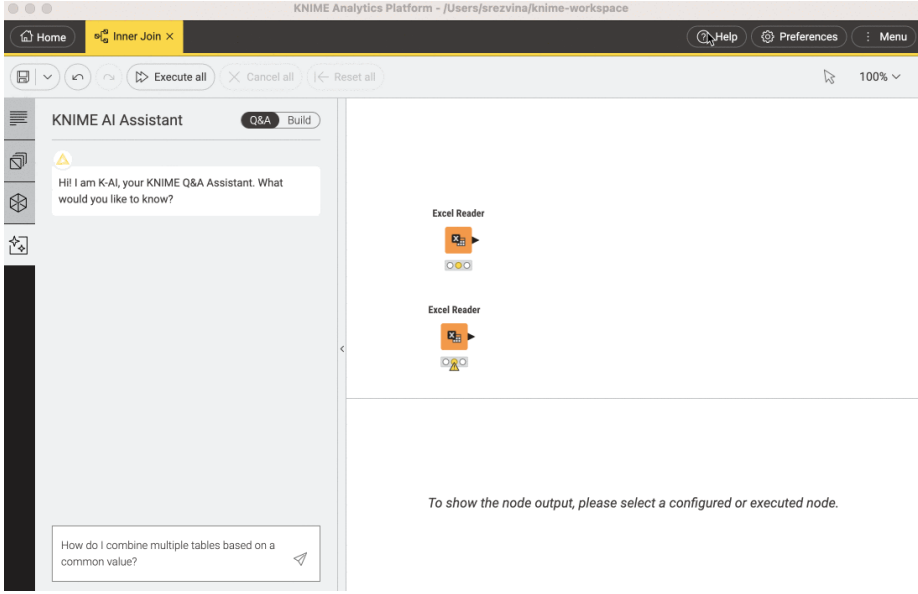
Protect Data, Monitor Consumption, Control Access

Learn and Get Assistance

…meet KAI, the KNIME AI

- Helps write code (Python, Expressions, …)
- Helps configure visualizations

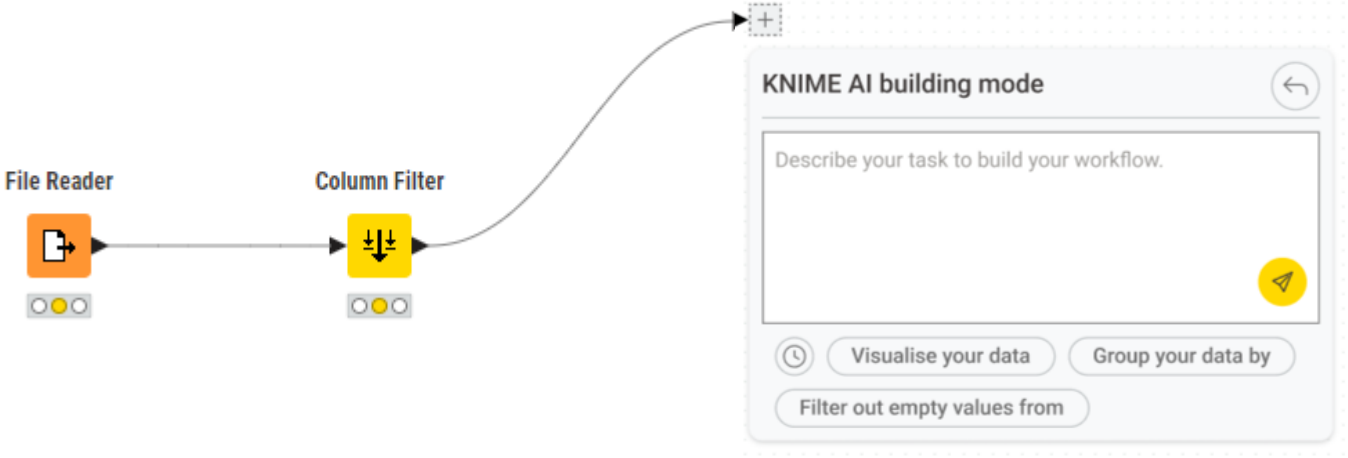




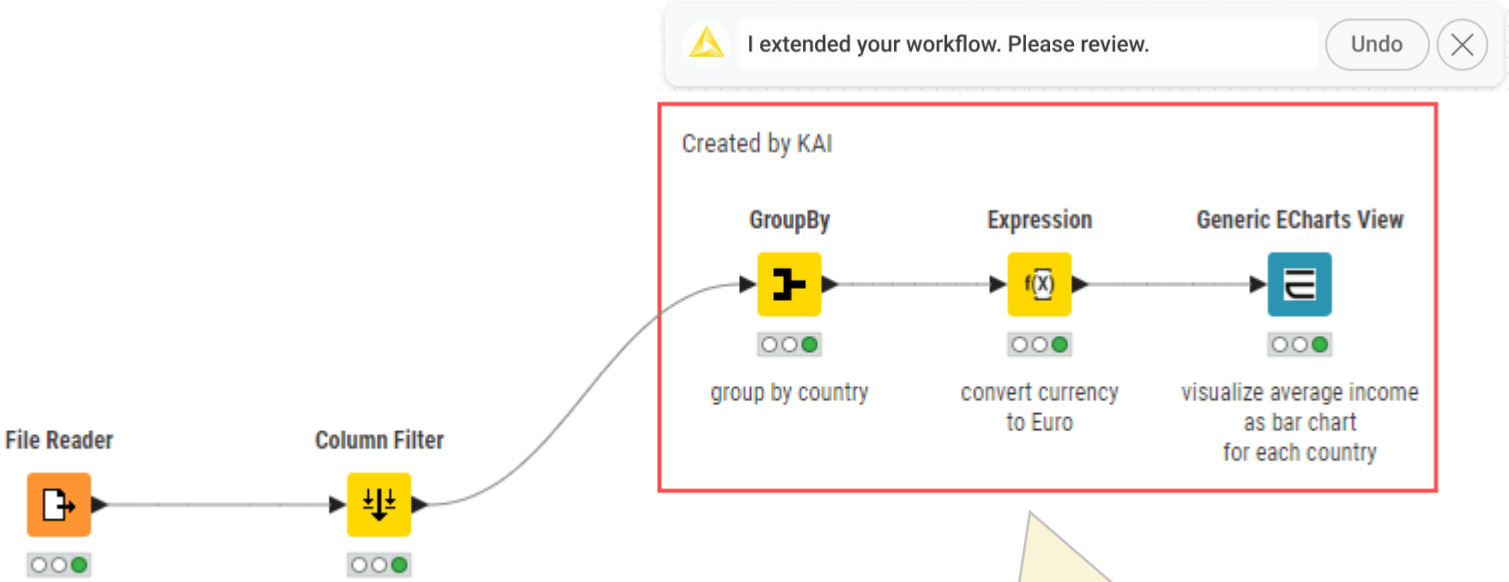
Delegate

…meet KAI, the KNIME AI

- Helps & collaborates when building workflows

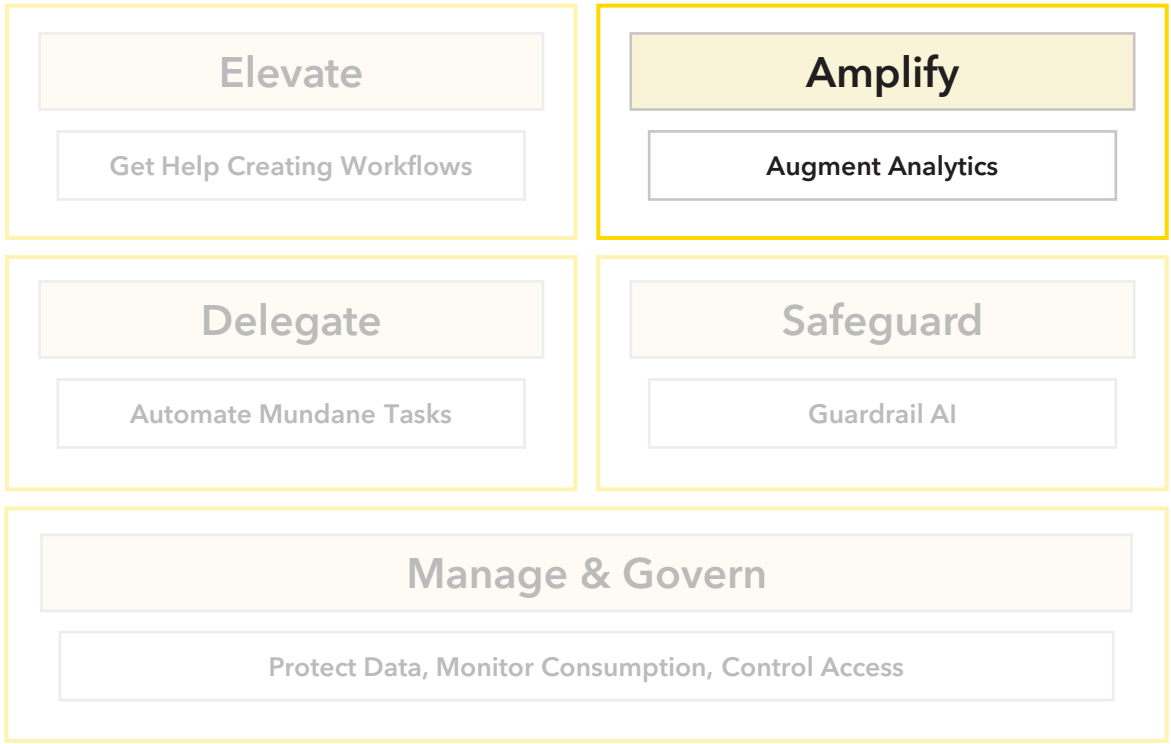


Delegate



Underneath the hood:

- Nodes are inserted by an AI 1-by-1
- configuration is defined via another AI



Augment Analytics

Open for Innovation KNIME | Hub

Search workflows, nodes and more...

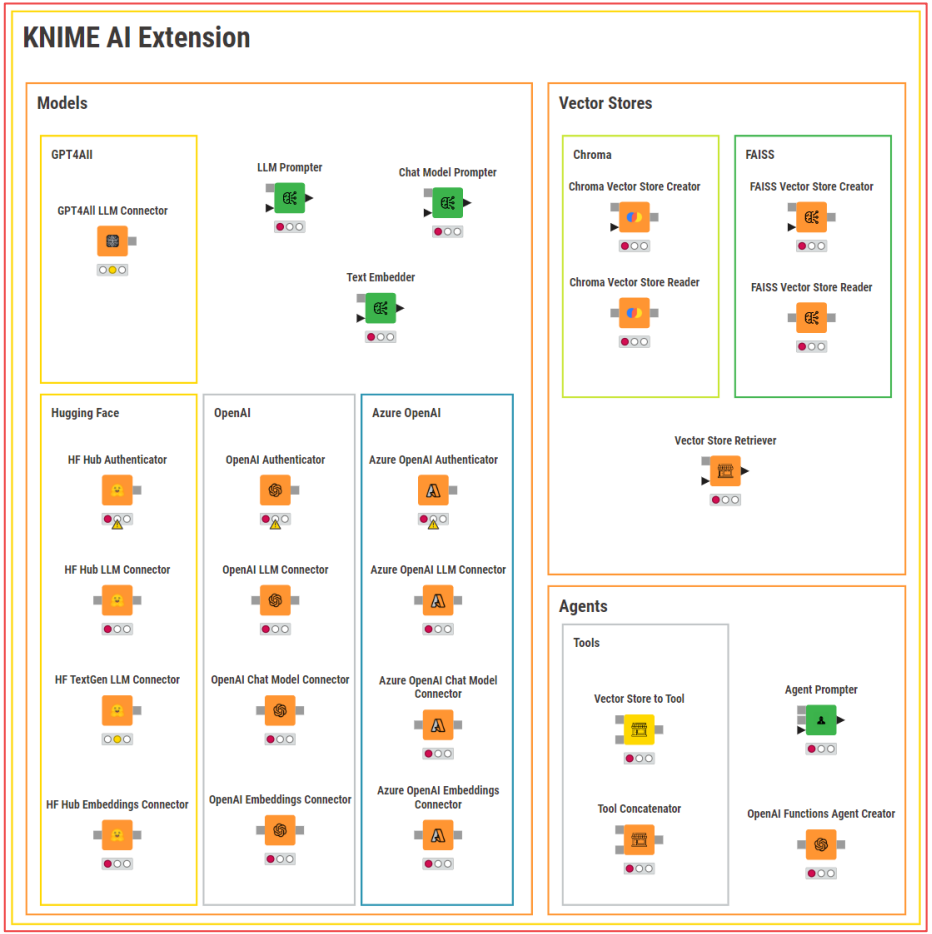
KNIME Community Hub > knime > Spaces > AI Extension Example Workflows

Public space

AI Extension Example Workflows

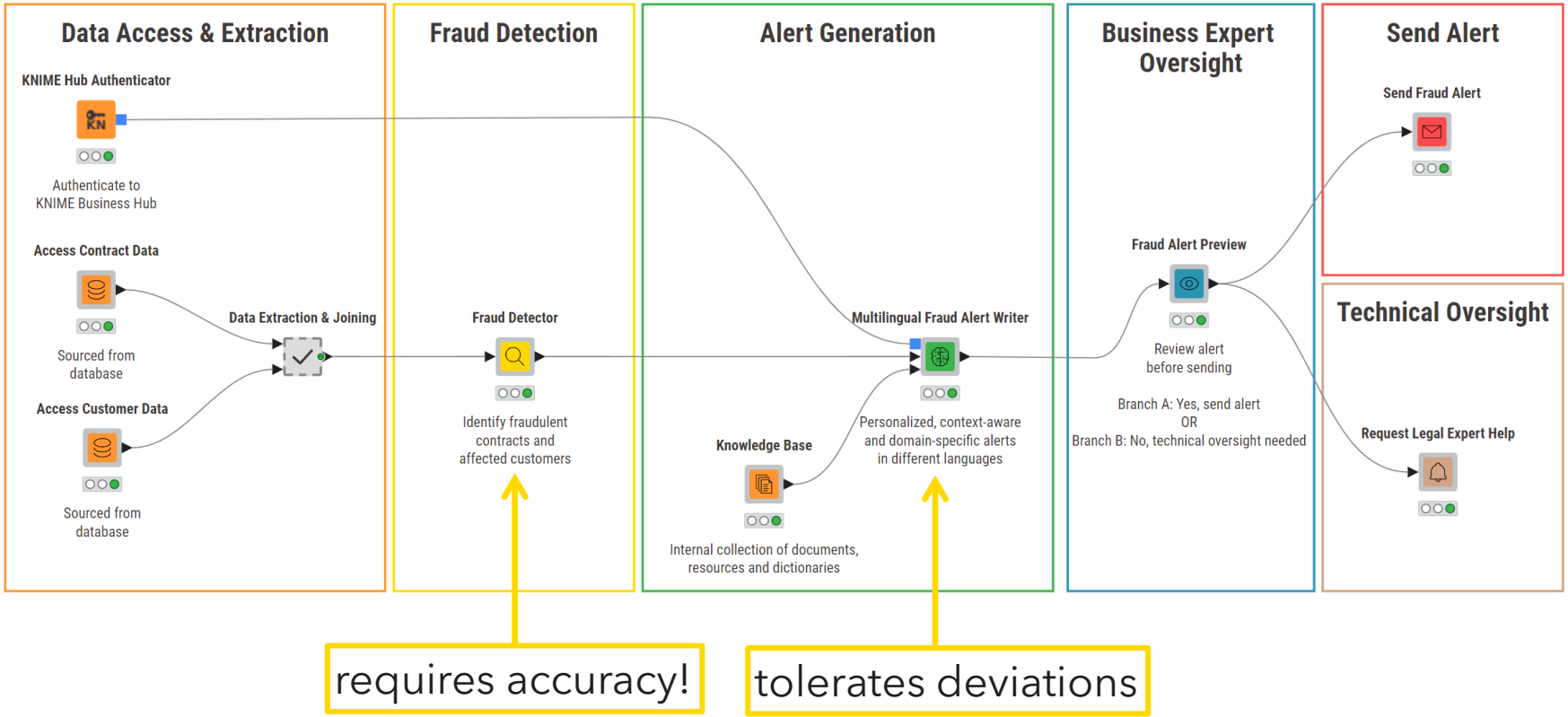
Home

- 1) Large Language Models
- 2) Chat Model
- 3) Vector Stores
- 4) Agents

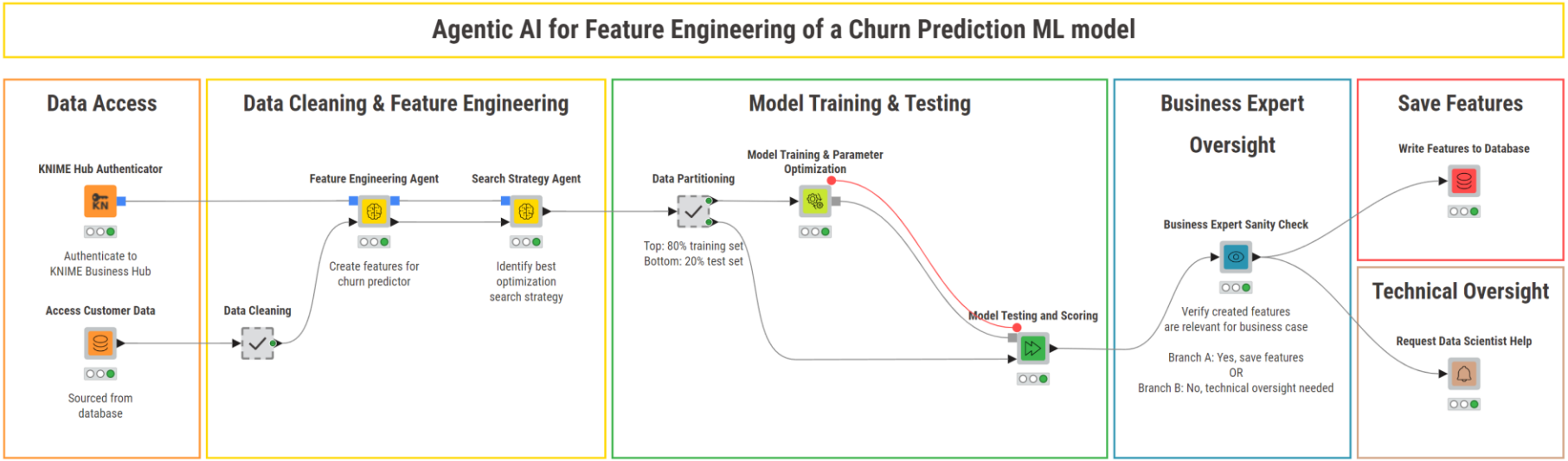


Open for Innovation KNIME

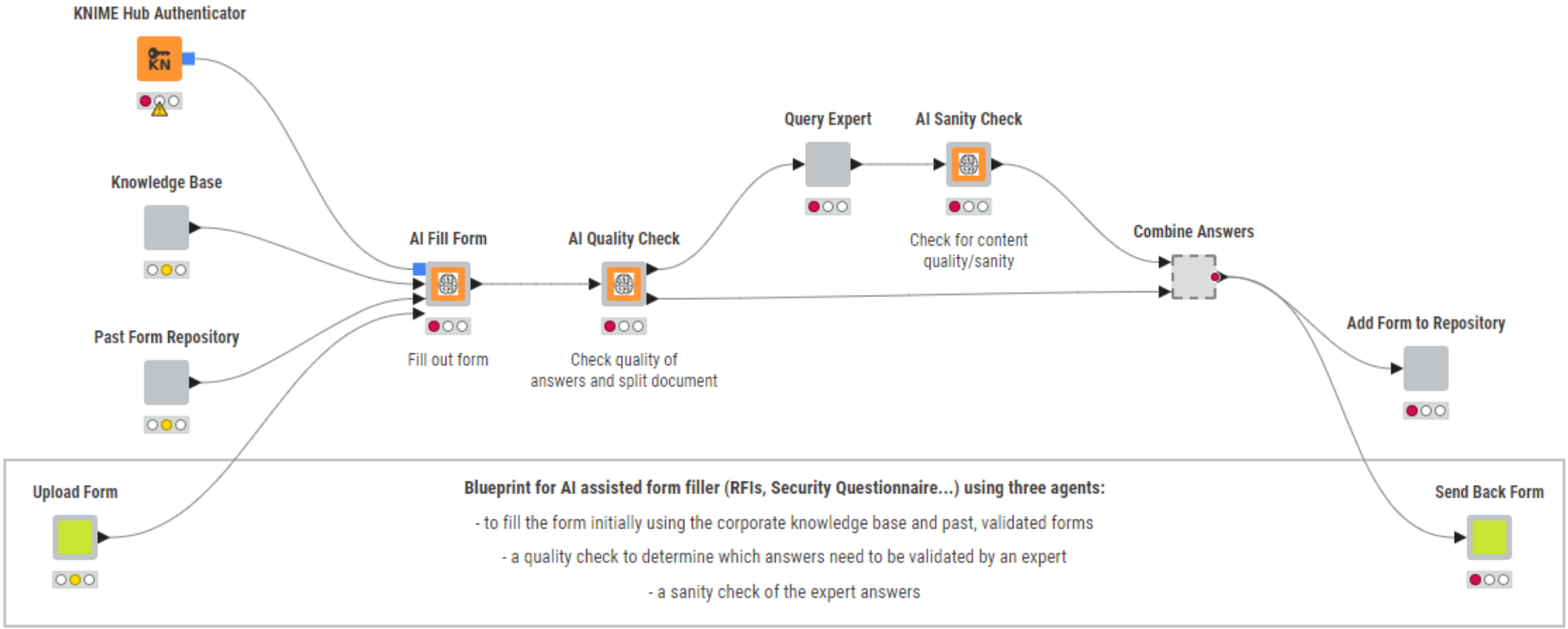
Augment Analytics: Fraud Detection & Alerting

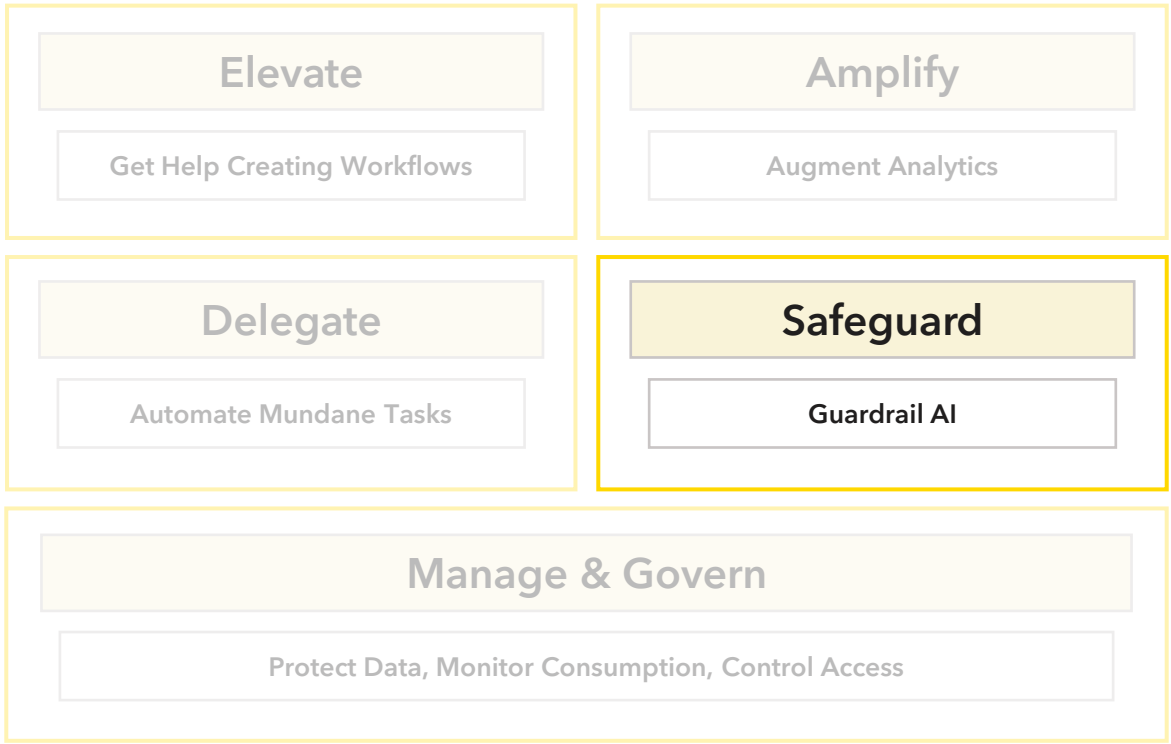


Augment Analytics with Data Science Expertise



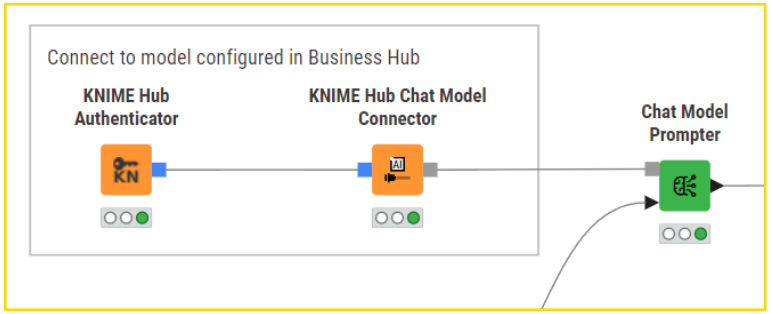
Intelligent Data Apps: Questionnaire Assistant (RFI, …)

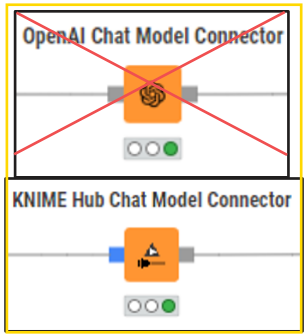


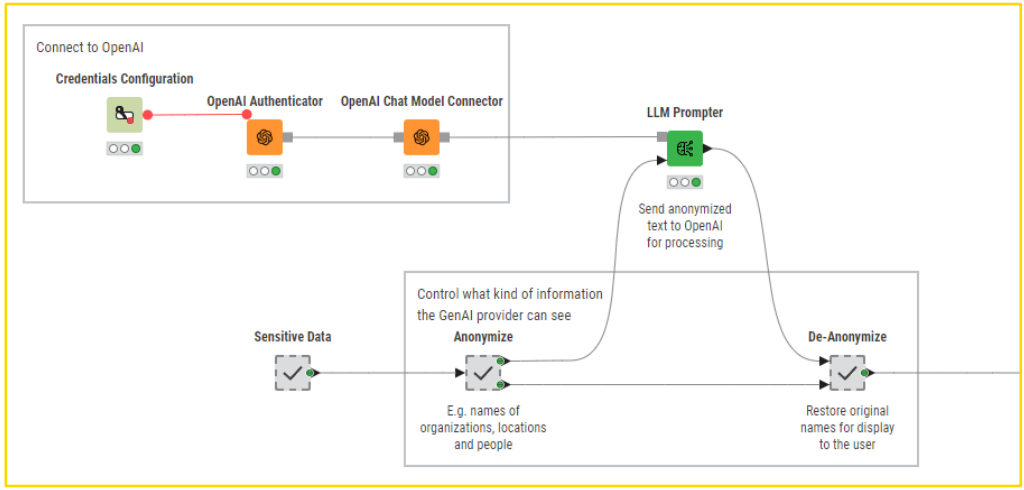


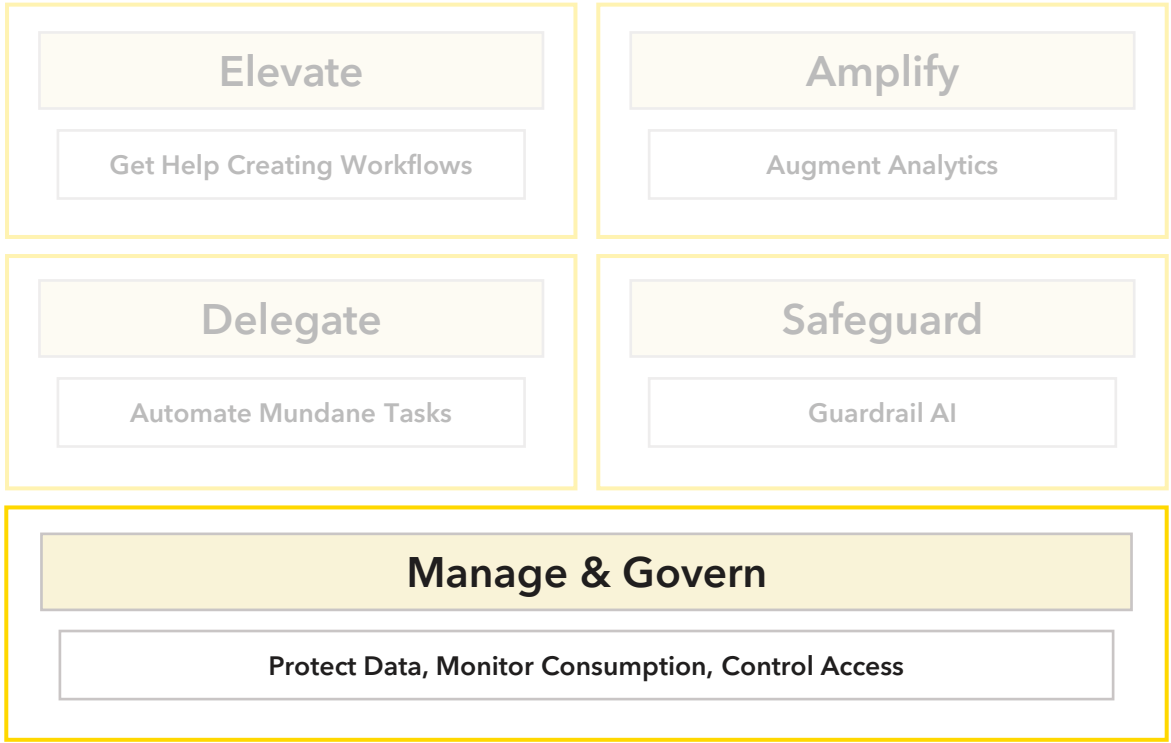
Safeguarding AI

- GenAI Gateway: forcing of internal use cases (workflows) to use approved AI (in conjunction with blacklisting “wild” access connectors)
- Configurable safeguards around AI services









Side Note: Governance in Data Science is Not New

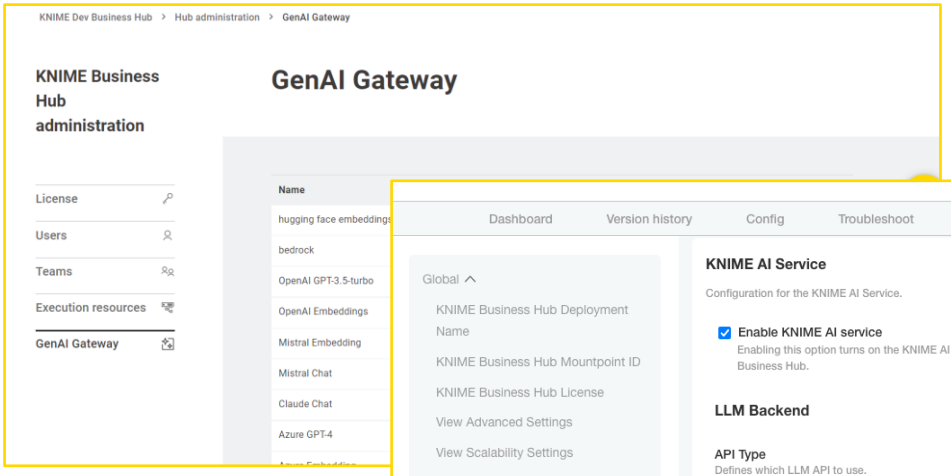
Take (& expand) governance tools proven with leading orgs

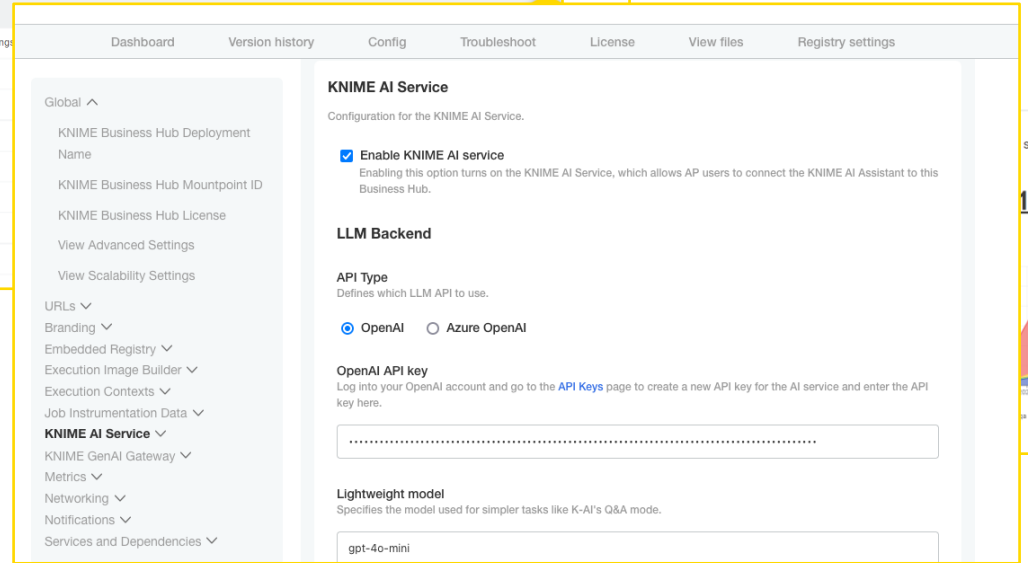
Govern data sent to models	Govern access to technology	Govern output & quality of models	Govern regulatory compliance
• KNIME workflows • KNIME components • KNIME Hub Permissioning • Anonymization Components • KNIME Hub GenAI Gateway • Presidio integration	• KNIME workflows • KNIME components • Trusted Extensions • Customization of KNIME Analytics Platform • KNIME Scripting Nodes • LLM integrations (Hugging face, etc)	• KNIME workflows • Explainable AI Components • Model Factory Framework • Continuous Deployment for Data Science (CDDS) extension • Giskard integration	• KNIME workflows • KNIME components • KNIME Hub Versioning • Continuous Deployment for Data Science (CDDS) extension • Metadata mapping • AI Workflow Explainer

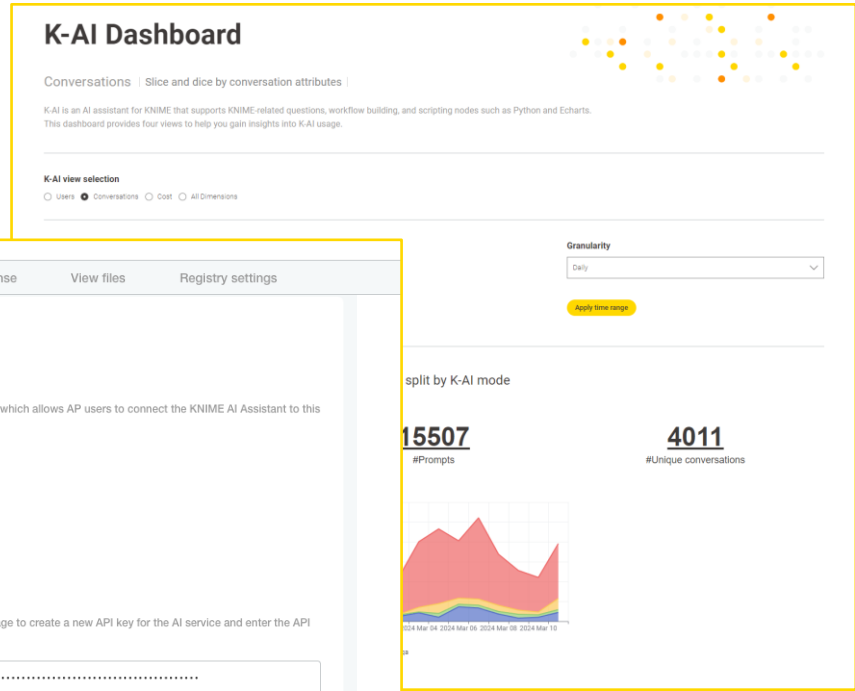
Controlling models we don't understand isn't new.
It's just now top of everybody's mind.

Managing and Governance of AI

- GenAI Gateway: routing of assistants to configurable local or cloud-based AI services
- Monitor & control consumption



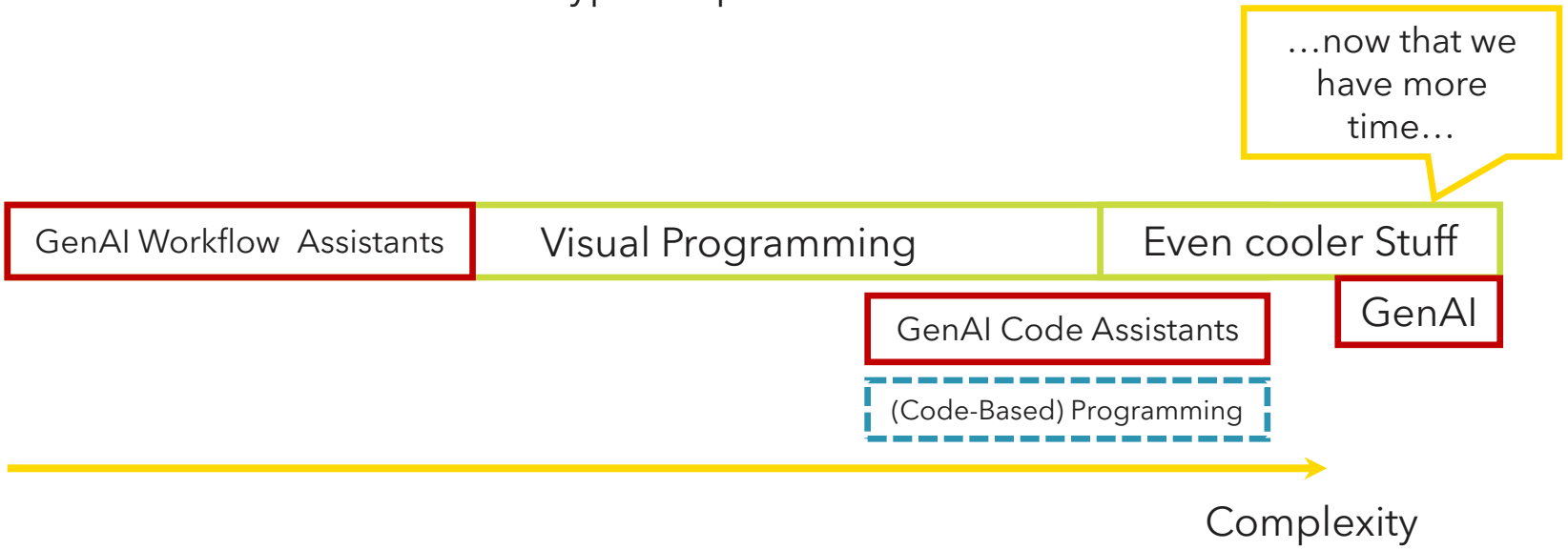




Quo Vadis?

Quo Vadis? Creating Workflows Together with AI

GenAI assisted workflows for all types of problems:



Quo Vadis: Intelligent Workflows

- Augment analytical workflows with reliable (when required) AI integrations
- Deploy reliable agents for other agents to use
- Deploy reliable agentic Data Apps to others in the organization

and, increasingly:

- Workflows as the collaboration language between humans and AI:
  - Explanations for AI reasoning (“this is how I reached the conclusion”)
  - Suggestions (“why don’t you try that?”)
  - Intelligent Expert Knowledge Base (“your colleagues would (not) do that.”)

Increasing support by assistants/agents – complete creation of complex workflows far out

Exciting Times ahead…

Trust AI with your Data – by using Workflows

Thank You!