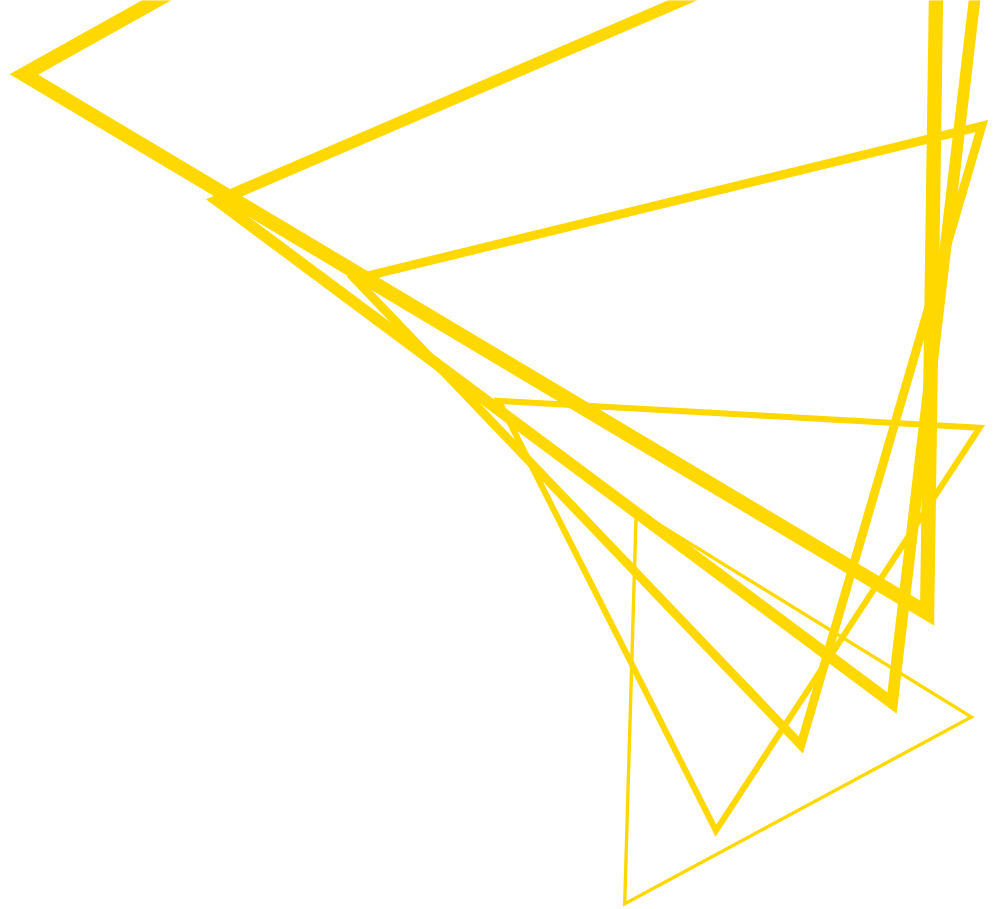


Data Science and GenAI

Michael Berthold

DataHop Singapore. December 10, 2024



Abstract geometric lines in the top right corner, consisting of several overlapping, thin white lines forming a complex, angular shape against the yellow background.

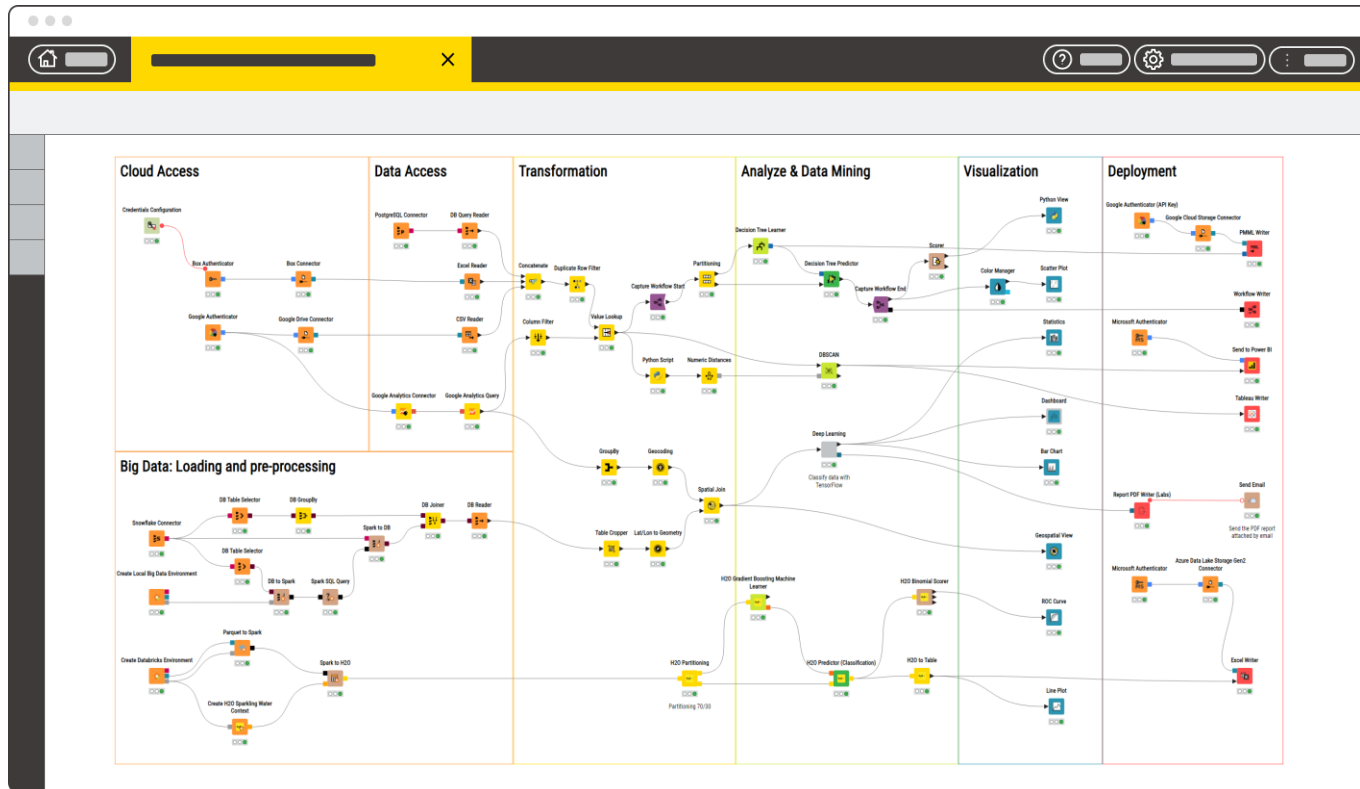
Setting the stage:

**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.

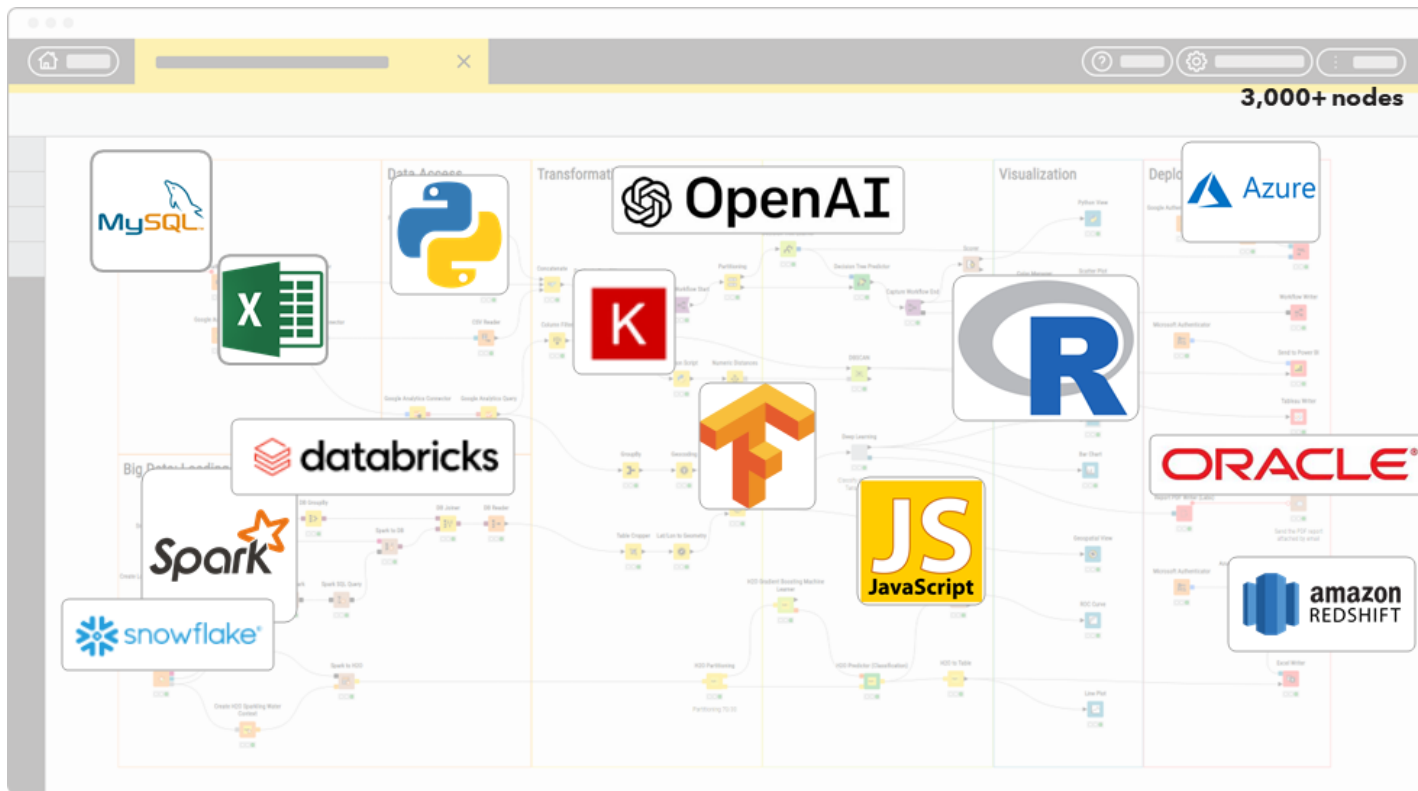
Visual Workflows for Data Science & AI

Provide techniques & capabilities to all types of users



Visual Workflows for Data Science & AI

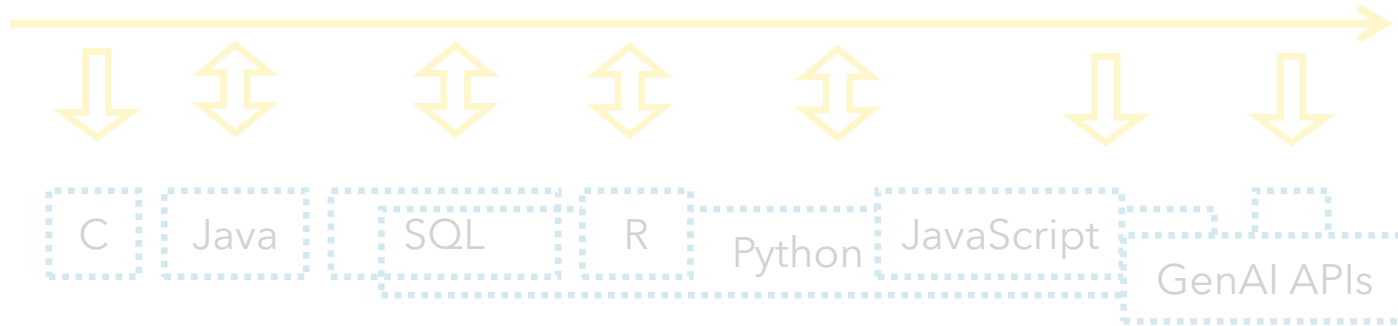
Abstract away unnecessary complexity and implementation details



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

Does it work?

“Instead of teaching students how to program for months, they can build their first workflow doing actual data work on day 2.”



Bob Peers • 1st
Chief Operating Officer at NN07
1mo • 🌐



Cloudless AI in Record Time: Deployed a Local LLM in 30 Minutes Flat 🕒

Using AI and LLMs (large language models) is available to virtually everyone these days via a web browser, but using it in production workflows, leveraging locally run, open source models seem to be the next frontier.

To that end I decided to see how easy, or difficult, it would be to get a local LLM up and running on my own laptop.

At this point I have a local, private, secure LLM running sentiment analysis for the total cost of precisely nothing. Amazing 🤩

#llm #knime #opensource #ai



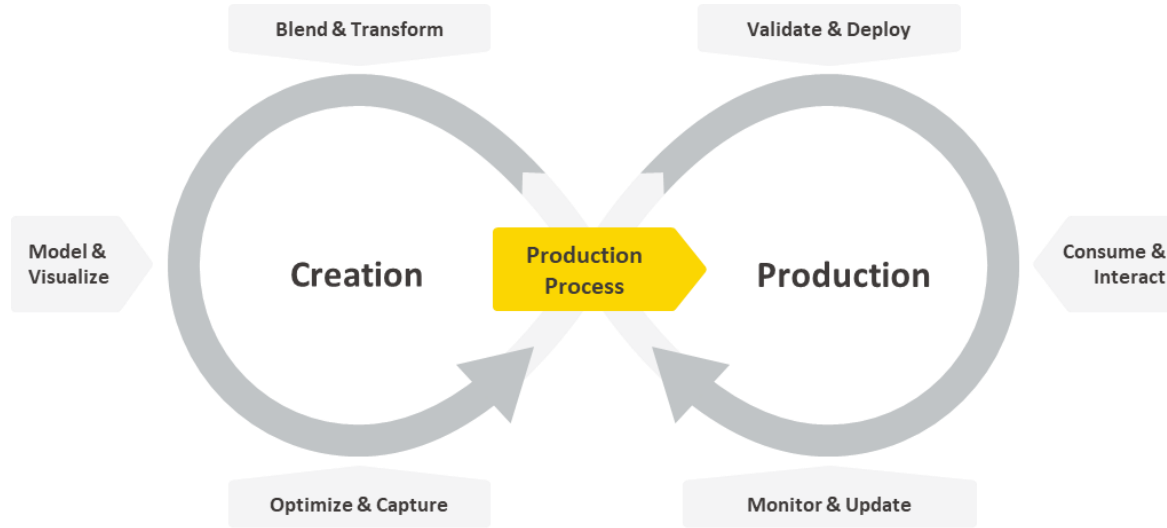
👤 You and 86 others

14 comments · 2 reposts

Abstract geometric lines in the top right corner, consisting of several overlapping, thin white lines forming a complex, angular shape against the yellow background.

**Quick Reminder:
KNIME Workflows & Software**

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...



Is AI out of control?

Enterprises lack oversight over technology use.

82%

of IT leaders already report security incidents due to AI¹

56%

of enterprises report shadow AI use¹

30%

of cloud spend is wasted due to lack of visibility & overprovisioning²

Can we (ever) trust AI?

- Create images
- Provide starting points for documents
- Control a character in a game

- Hallucinations ok
 - Fake information tolerable
- ⇒ Trust AI to be “good enough”

- Write code
- Create legal documents
- Control a nuclear power plant

- Hallucinations not ok
 - Fake outputs unacceptable
- ⇒ Careful supervision/proofing required!

Can we (ever) trust AI when working with data?

- Exploratory Data Analysis
- Creating hypotheses
- Providing first insights
- Hallucinations ok
- Fake information tolerable
- ⇒ Trust AI to be “good enough”
- Create tax reports
- Generate business forecasts
- Train predictive models without bias
- Hallucinations not ok
- Fake outputs unacceptable
- ⇒ Careful supervision/proofing required!

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

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

Abstract geometric lines in the top right corner, consisting of several overlapping, thin white lines forming various triangular and polygonal shapes against a yellow background.

Trust AI with Your Data

Leverage AI to help or do data work



Elevate

Get Help Creating Workflows



Delegate

Automate Mundane Tasks

Build new solutions with AI technology



Amplify

Augment Analytics



Safeguard

Guardrail AI



Manage & Govern

Protect Data, Monitor Consumption, Control Access

KNIME and AI



Leverage AI to help or do data work



Elevate

Get Help Creating Workflows



Delegate

Automate Mundane Tasks

Build new solutions with AI technology



Amplify

Augment Analytics



Safeguard

Guardrail AI



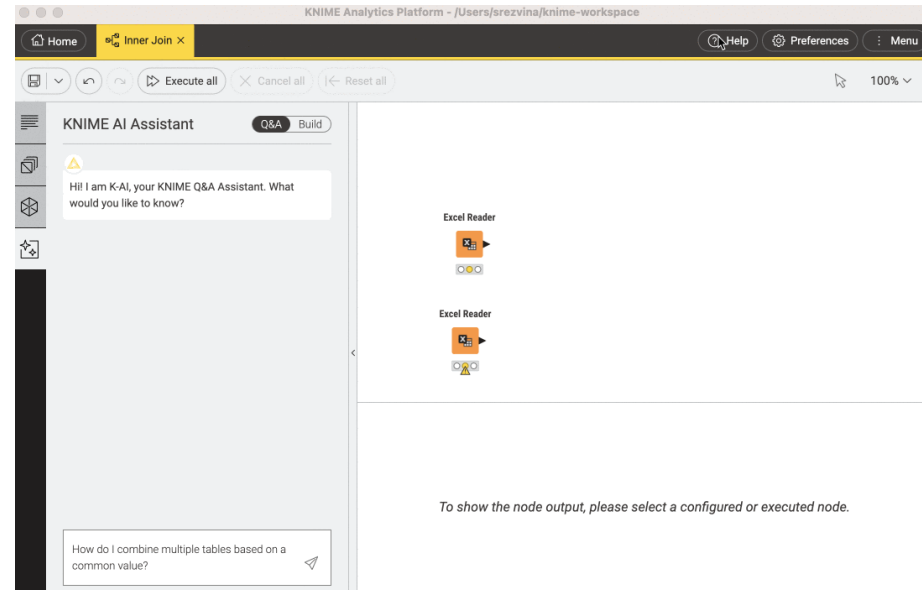
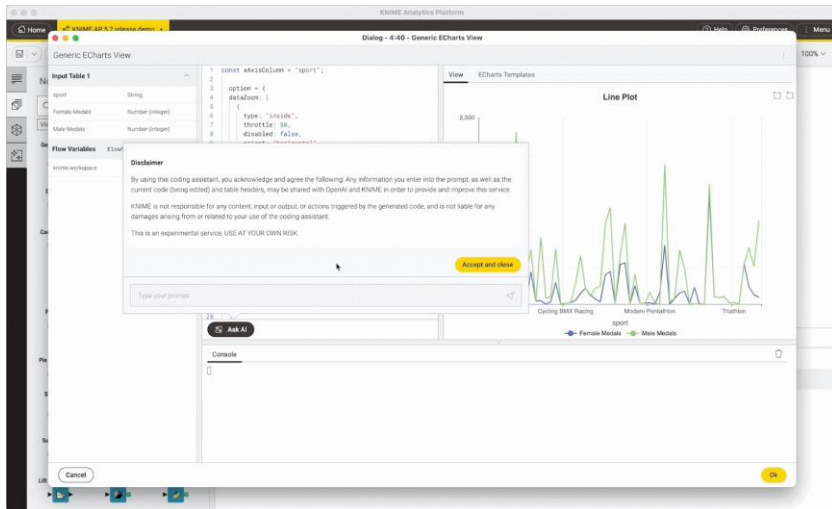
Manage & Govern

Protect Data, Monitor Consumption, Control Access

Learn and Get Assistance

...meet KAI, the KNIME AI

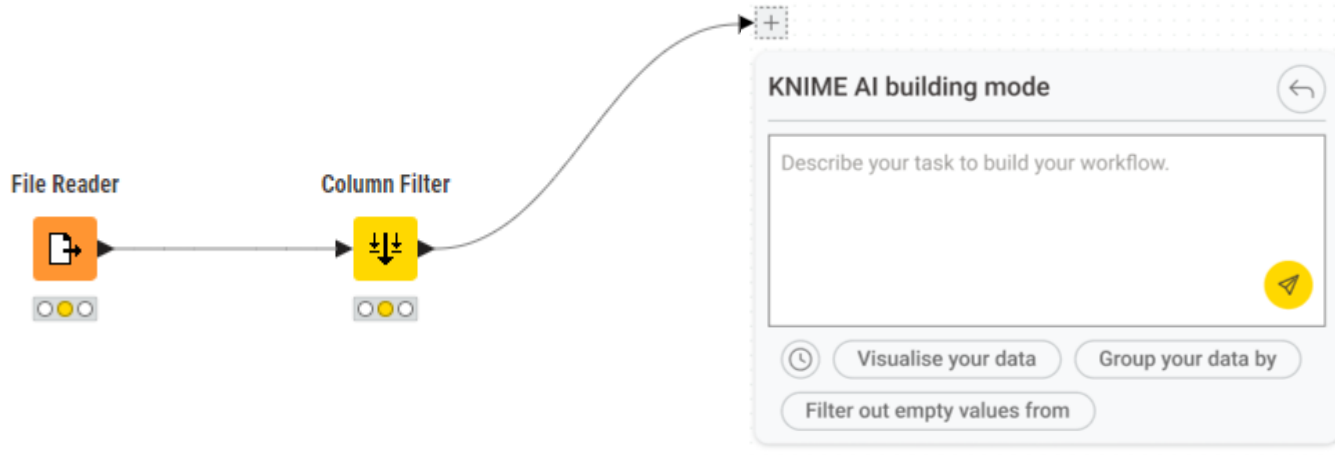
- Helps write code (Python, Expressions, ...)
- Helps configure visualizations



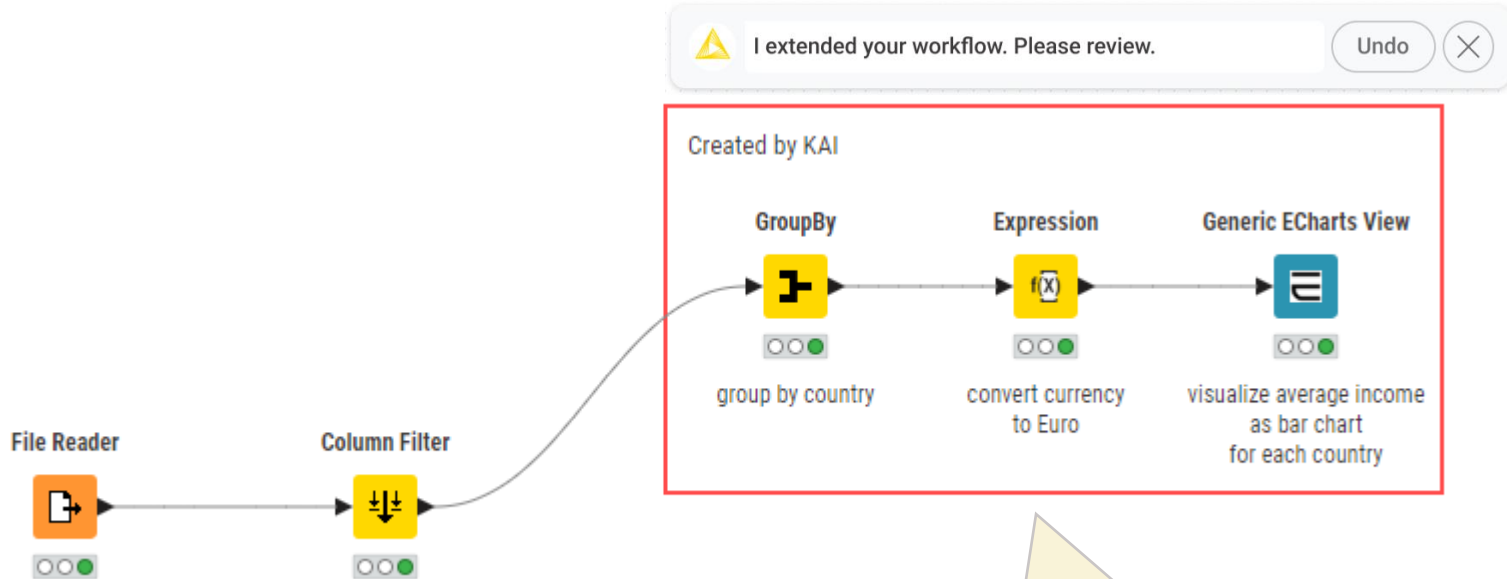
Delegate

...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

Leverage AI to help or do data work



Elevate

Get Help Creating Workflows



Delegate

Automate Mundane Tasks

Build new solutions with AI technology



Amplify

Augment Analytics



Safeguard

Guardrail AI



Manage & Govern

Protect Data, Monitor Consumption, Control Access

Augment Analytics

AI Extension Example Workflows

Home

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

KNIME AI Extension

Models

GPT4All

GPT4All LLM Connector



LLM Prompter



Chat Model Prompter



Text Embedder



Hugging Face

HF Hub Authenticator



HF Hub LLM Connector



HF TextGen LLM Connector



HF Hub Embeddings Connector



OpenAI

OpenAI Authenticator



OpenAI LLM Connector



OpenAI Chat Model Connector



OpenAI Embeddings Connector



Azure OpenAI

Azure OpenAI Authenticator



Azure OpenAI LLM Connector



Azure OpenAI Chat Model Connector



Azure OpenAI Embeddings Connector



Vector Stores

Chroma

Chroma Vector Store Creator



Chroma Vector Store Reader



FAISS

FAISS Vector Store Creator



FAISS Vector Store Reader



Vector Store Retriever



Agents

Tools

Vector Store to Tool



Tool Concatenator



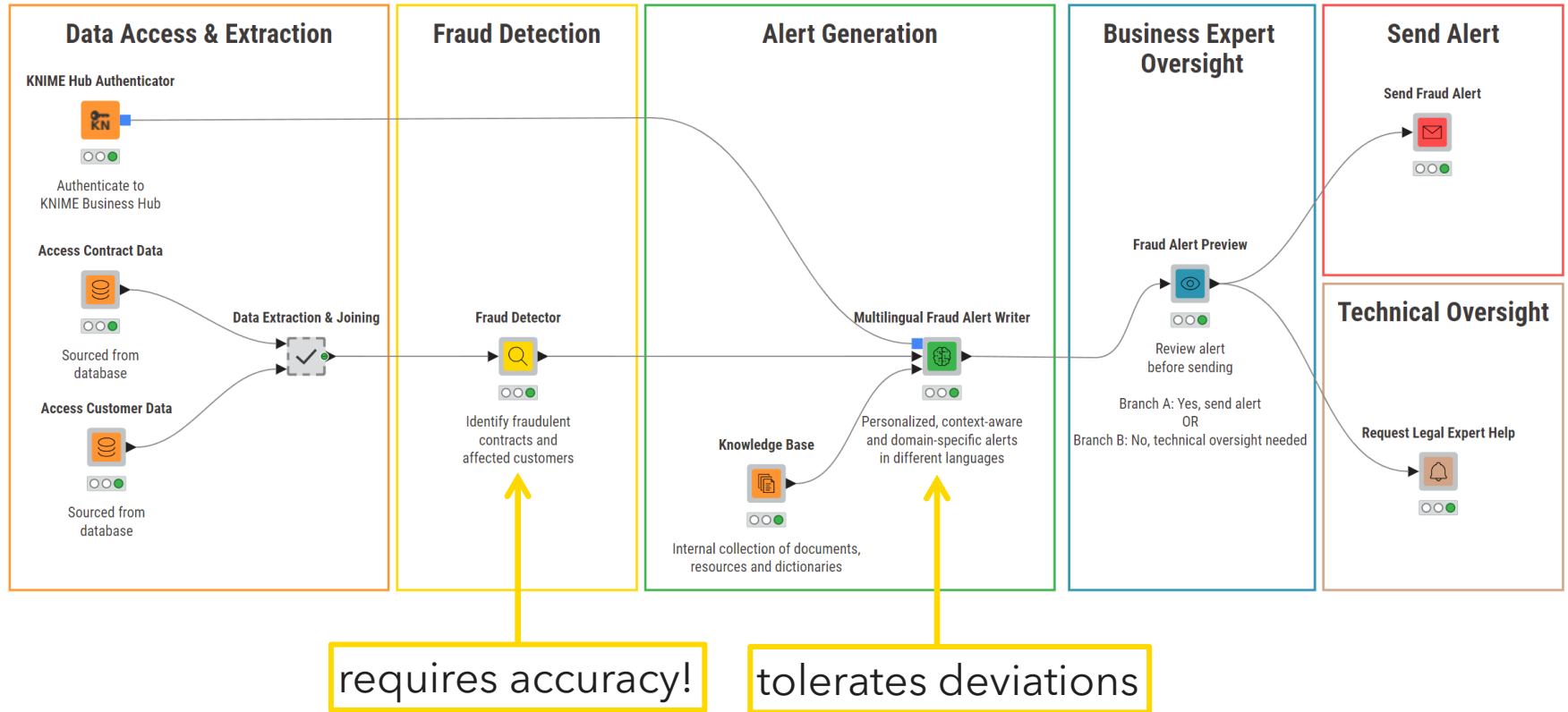
Agent Prompter



OpenAI Functions Agent Creator

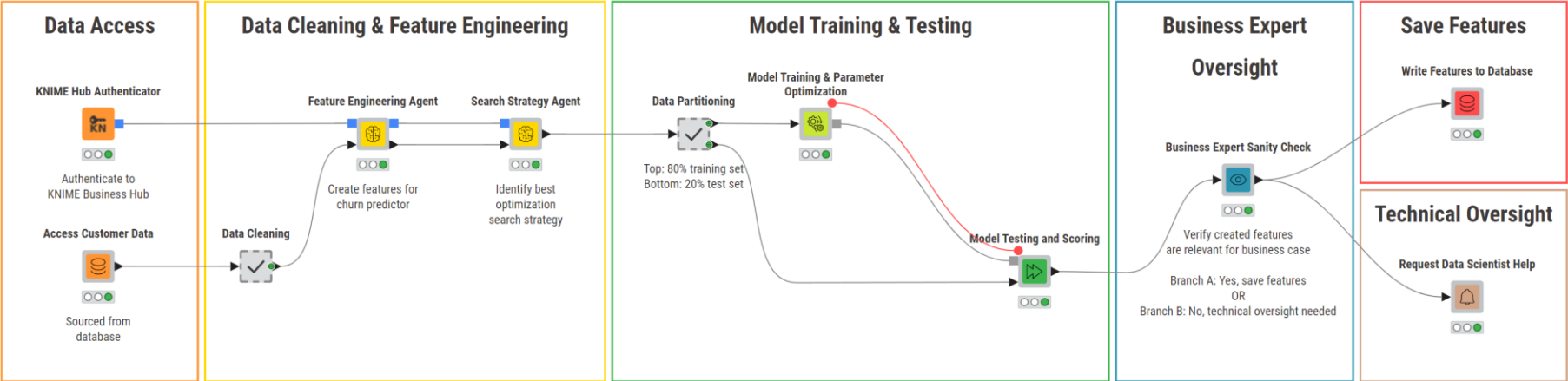


Augment Analytics: Fraud Detection & Alerting

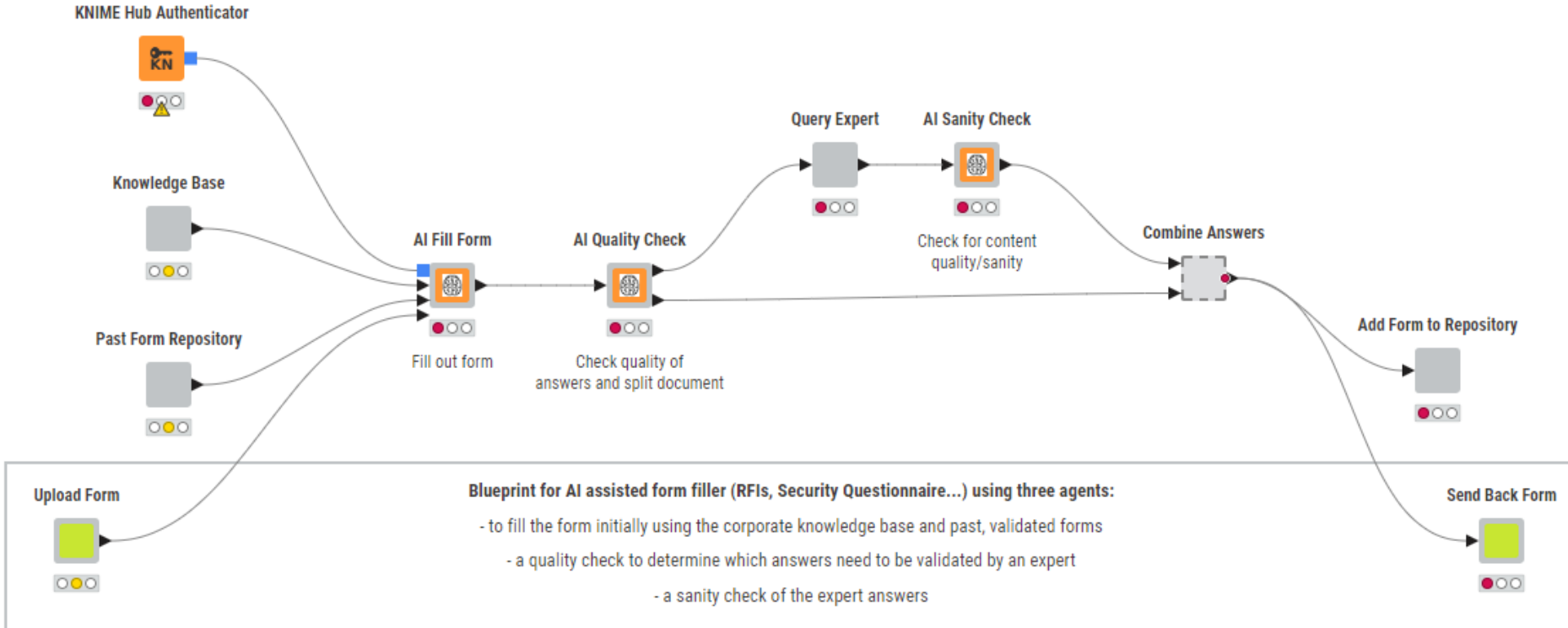


Augment Analytics with Data Science Expertise

Agentic AI for Feature Engineering of a Churn Prediction ML model



Intelligent Data Apps: Questionnaire Assistant (RFI, ...)



Leverage AI to help or do data work



Elevate

Get Help Creating Workflows



Delegate

Automate Mundane Tasks

Build new solutions with AI technology



Amplify

Augment Analytics



Safeguard

Guardrail AI

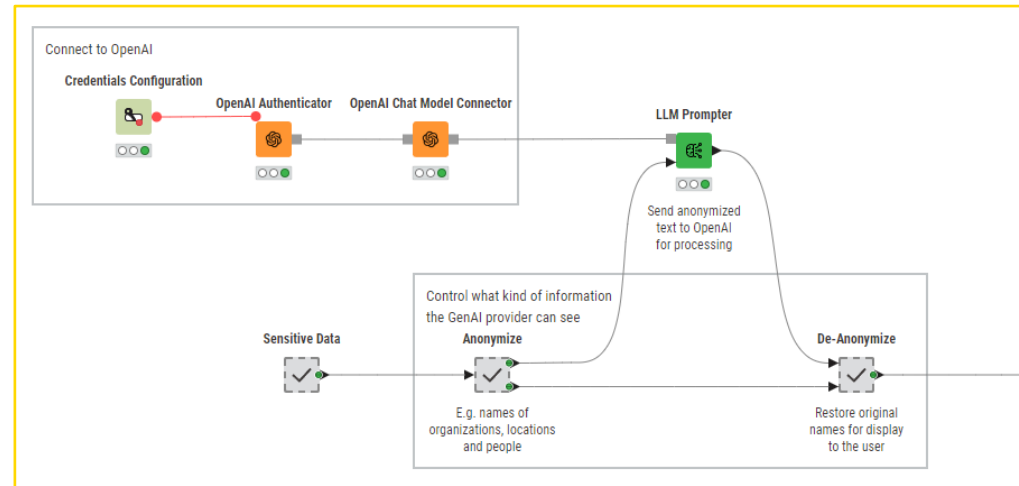
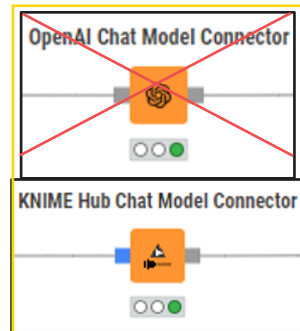
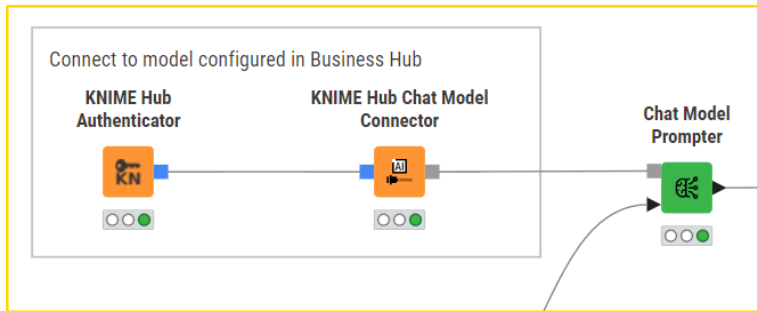


Manage & Govern

Protect Data, Monitor Consumption, Control Access

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



Leverage AI to help or do data work



Elevate

Get Help Creating Workflows



Delegate

Automate Mundane Tasks

Build new solutions with AI technology



Amplify

Augment Analytics



Safeguard

Guardrail AI



Manage & Govern

Protect Data, Monitor Consumption, Control Access

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 |
|--|--|---|--|
| <ul style="list-style-type: none">• KNIME workflows• KNIME components• KNIME Hub• Permissioning• Anonymization• Components• KNIME Hub GenAI Gateway• Presidio integration | <ul style="list-style-type: none">• KNIME workflows• KNIME components• Trusted Extensions• Customization of KNIME Analytics Platform• KNIME Scripting Nodes• LLM integrations (Hugging face, etc) | <ul style="list-style-type: none">• KNIME workflows• Explainable AI Components• Model Factory Framework• Continuous Deployment for Data Science (CDDS) extension• Giskard integration | <ul style="list-style-type: none">• 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

KNIME Dev Business Hub > Hub administration > GenAI Gateway

KNIME Business Hub administration

License

Users

Teams

Execution resources

GenAI Gateway

GenAI Gateway

| Name |
|-------------------------|
| hugging face embeddings |
| bedrock |
| OpenAI GPT-3.5-turbo |
| OpenAI Embeddings |
| Mistral Embedding |
| Mistral Chat |
| Claude Chat |
| Azure GPT-4 |

K-AI Dashboard

Conversations | Slice and dice by conversation attributes |

K-AI is an AI assistant for KNIME that supports KNIME-related questions, workflow building, and scripting nodes such as Python and Echarts. This dashboard provides four views to help you gain insights into K-AI usage.

K-AI view selection

Users Conversations Cost All Dimensions

Granularity

Daily

Apply time range

split by K-AI mode

15507
#Prompts

4011
#Unique conversations

2024 Mar 04 2024 Mar 06 2024 Mar 08 2024 Mar 10

KNIME AI Service

Configuration for the KNIME AI Service.

Enable KNIME AI service
Enabling this option turns on the KNIME AI Service, which allows AP users to connect the KNIME AI Assistant to this Business Hub.

LLM Backend

API Type
Defines which LLM API to use.

OpenAI Azure OpenAI

OpenAI API key
Log into your OpenAI account and go to the [API Keys](#) page to create a new API key for the AI service and enter the API key here.

Lightweight model
Specifies the model used for simpler tasks like K-AI's Q&A mode.

Outlook



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.”)

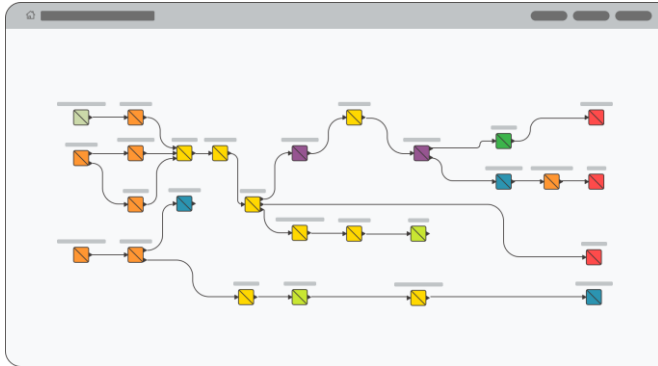
KNIME Software:

Balance Innovation & Control

Increase Innovation

KNIME Analytics Platform

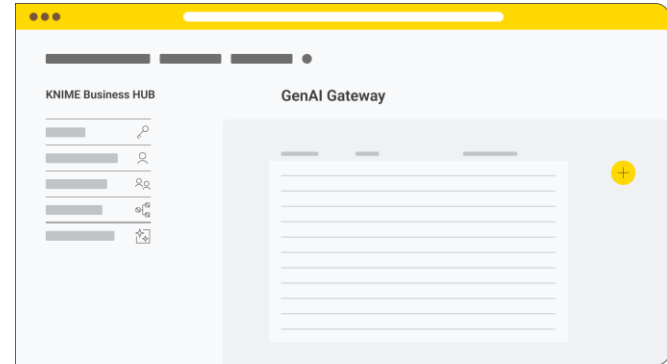
Bring the power of intuitive workflows to your entire organization



- Build & execute workflows of any complexity
- 1000s of blueprints & workflows for a fast start
- AI assistant (K-AI) for onboarding & high efficiency

KNIME Business Hub

Safeguard data work at scale to accelerate innovation



- Govern centrally, workflows & AI models
- Ensure safe automation & deployment
- Roll out best practices, globally

Reduce Risk

Exciting Times ahead...

Use Workflows so you can Trust AI with your Data

Thank You!