



DataHop Boston



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 + Workflow Doctor are available during all breaks

The top right corner of the slide features several overlapping, thin white lines that form abstract, geometric shapes, possibly representing data flow or network connections.

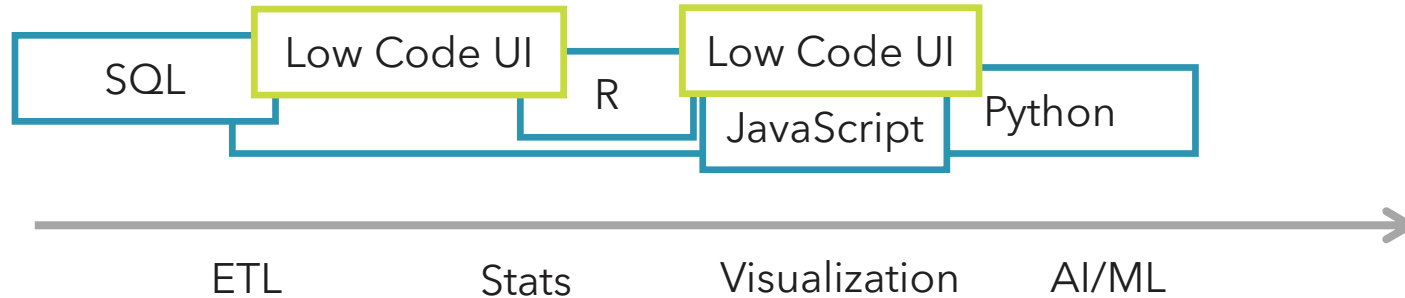
The KNIME stance 2022 and before:

**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)

An Intuitive Guide to Integrate SQL and Python for Data Science

Learn to master MySQL connector, a Python library that enables to interact with MYSQL database



Eugenia Anello · Follow

Published in Towards Data Science · 10 min read · 1 day ago

Combining R And Python In Jupyter Notebook: A Comprehensive Guide

Python

June 6, 2024 · Drummer Charlotte

When it comes to data analysis and scientific computing, two programming languages stand out: R and Python. R is renowned for its statistical capabilities,

Developing Interactive and Insightful Dashboards with Spark and Plotly Dash

Interactive Large-Scale Data Visualization for Web applications in Python



Yu Huang, M.D., M.S. in CS · Follow

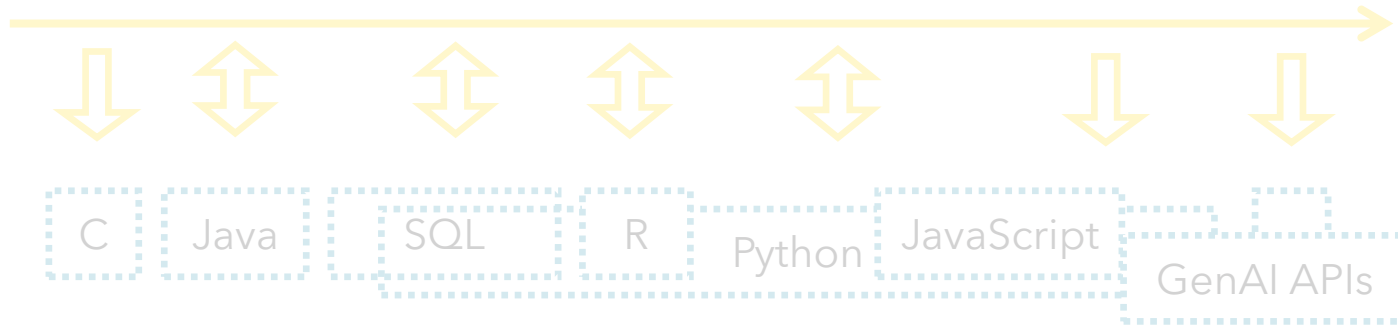
Published in Towards Data Science · 11 min read · Jun 21, 2023

- 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



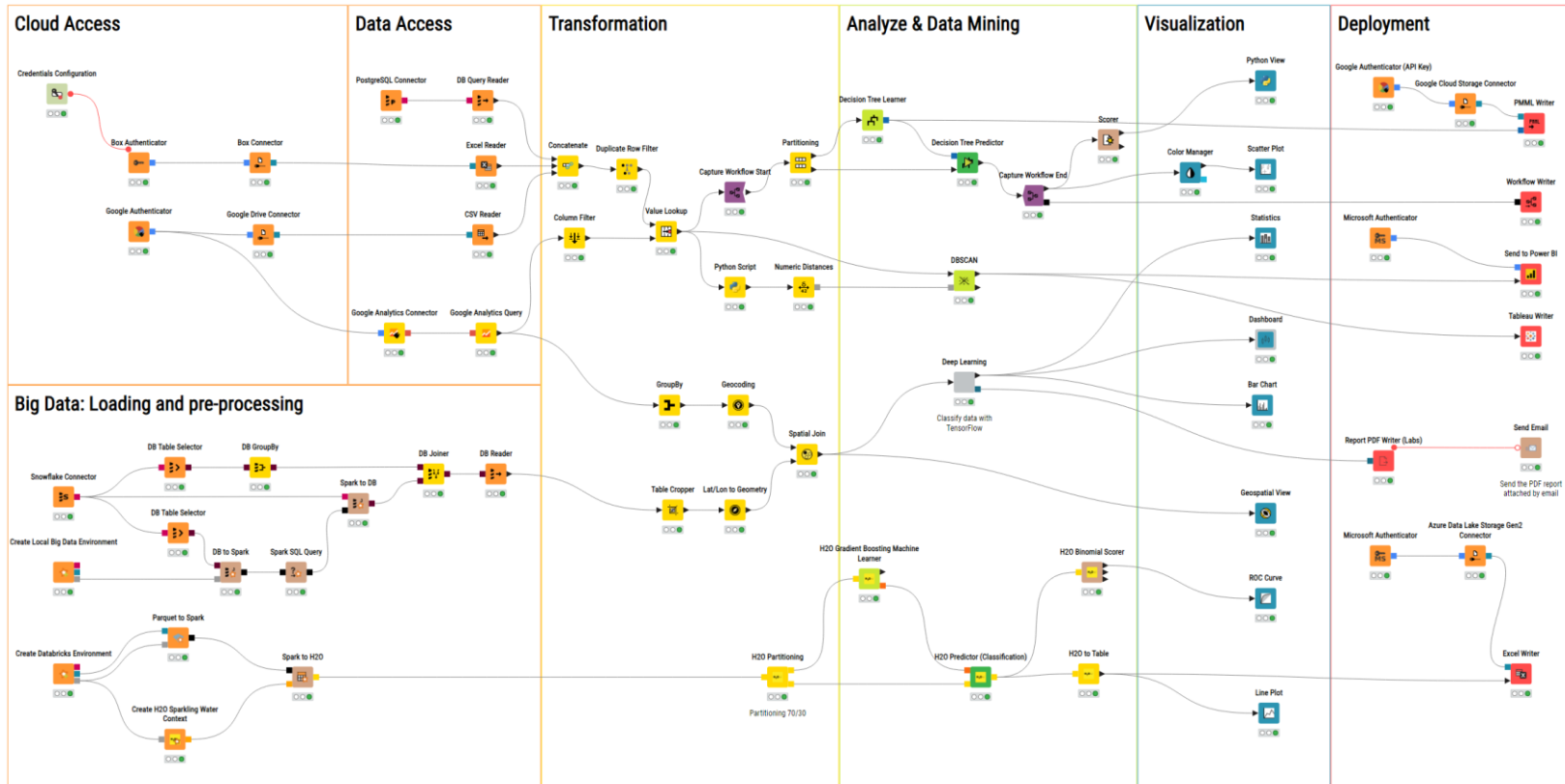
- Visual workflows **are** the programming language
- abstracts implementations in various technologies
- allows to embed code snippets
- Governed access to GenAI



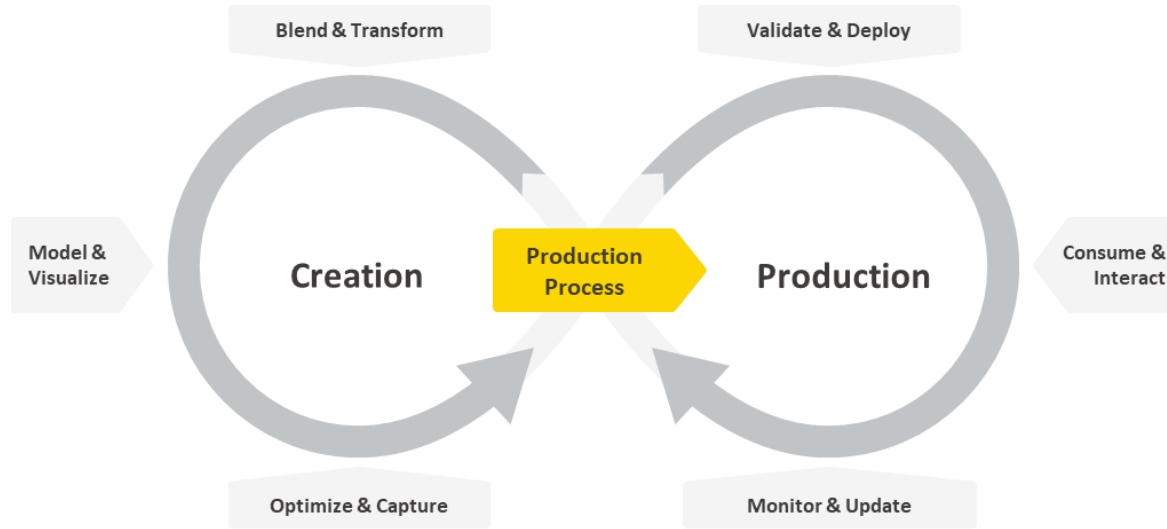
**Quick Reminder:
KNIME Workflows & Software**

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

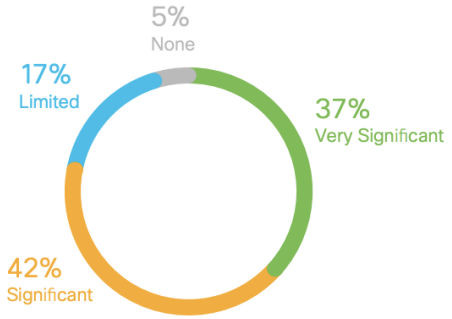
KNIME Community Hub		KNIME Business Hub		
Team Collaboration		Governance and Deployment at Scale		
Shared Repository	Data Apps & API Services	Automation & Orchestration	Continuous Delivery of Data Science	AI Governance

Enter: GenAI...



GenAI: Opportunities and Risks

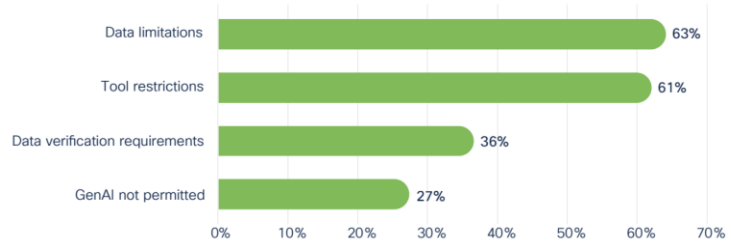
79% perceive significant value from GenAI adoption¹



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

Positive impact is blocked...

27% of organization have an outright ban on GenAI usage¹



- 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 top right corner of the slide features several overlapping, thin white lines that form abstract, geometric shapes, possibly representing data flow or network connections.

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

Elevate

Get Help Creating Workflows

Amplify

Augment Analytics

Delegate

Automate Mundane Tasks

Safeguard

Guardrail AI

Manage & Govern

Protect Data, Monitor Consumption, Control Access

KNIME and AI



Elevate

Get Help Creating Workflows

Amplify

Augment Analytics

Delegate

Automate Mundane Tasks

Safeguard

Guardrail AI

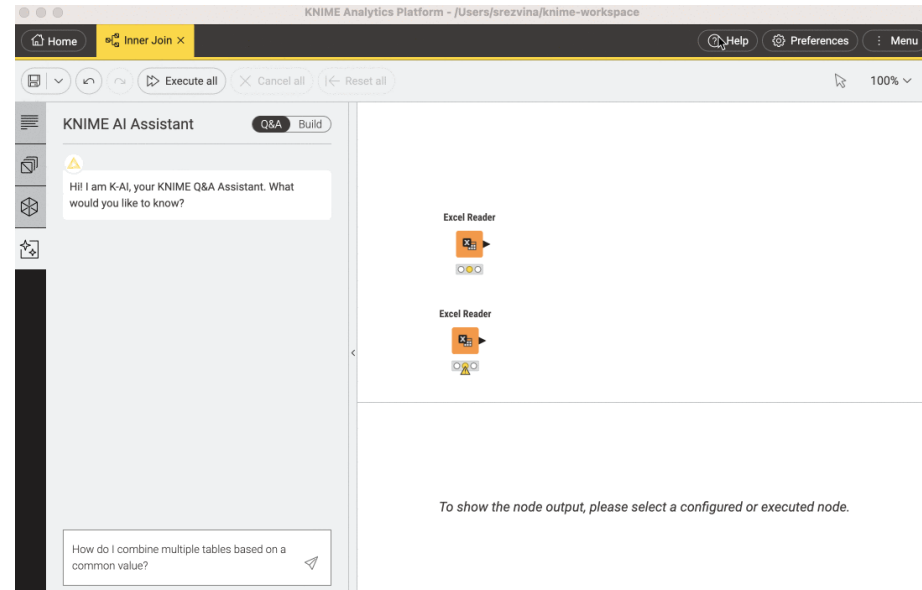
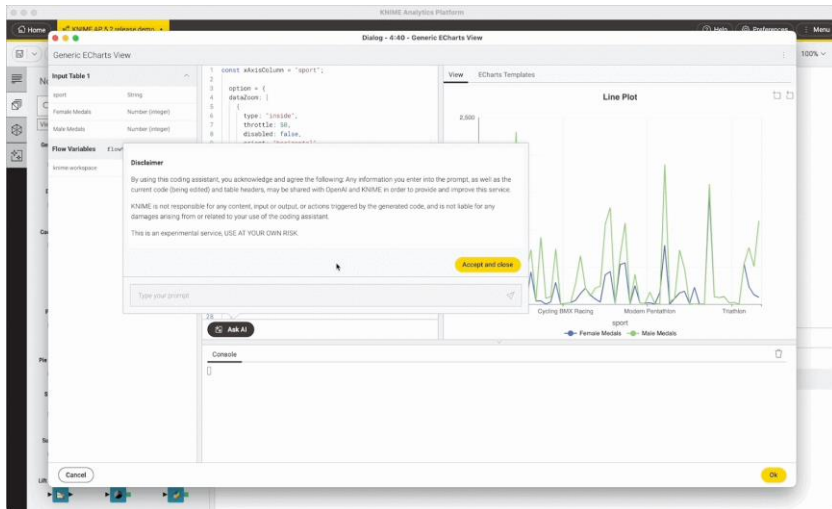
Manage & Govern

Protect Data, Monitor Consumption, Control Access

Learn and Get Assistance

...meet KAI, the KNIME AI

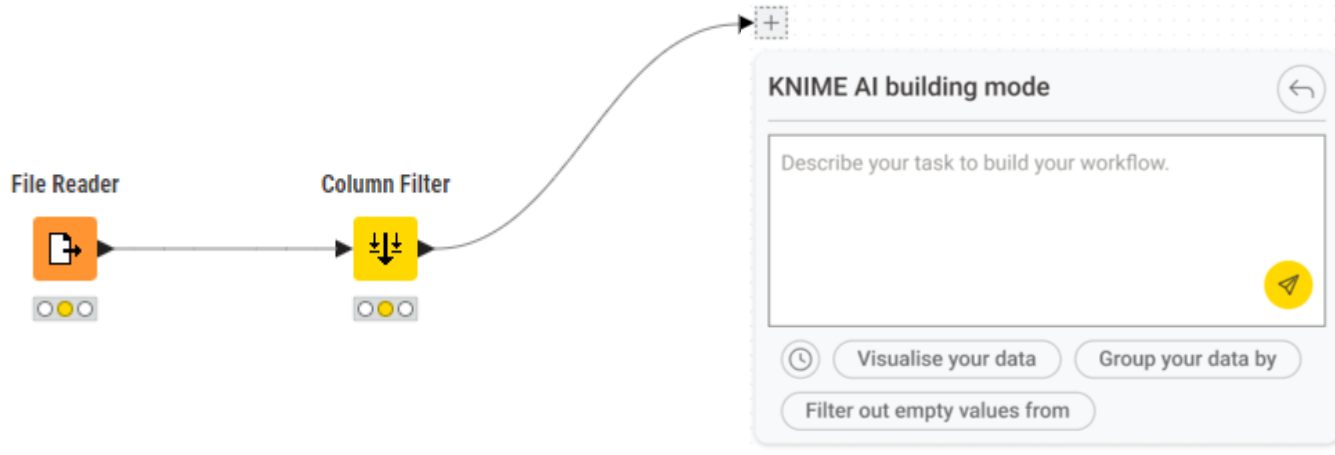
- Helps write code (Python, Expressions, ...)
- Helps configure visualizations



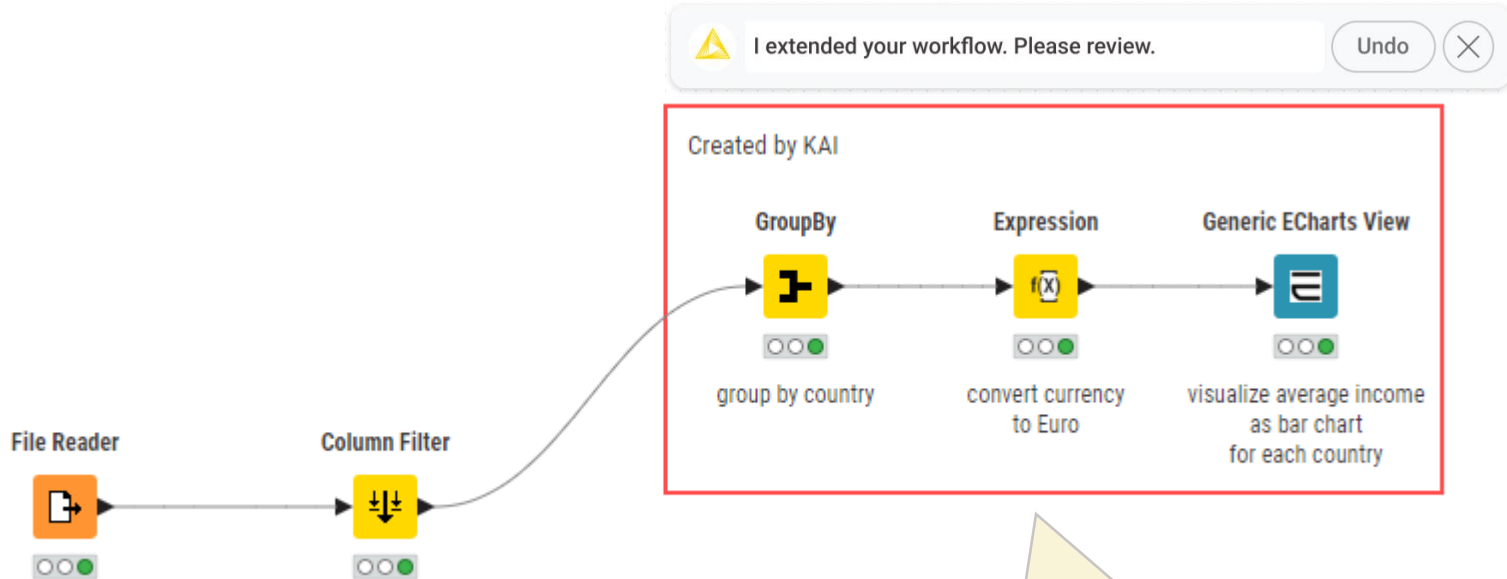
Delegate

...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

Elevate

Get Help Creating Workflows

Amplify

Augment Analytics

Delegate

Automate Mundane Tasks

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



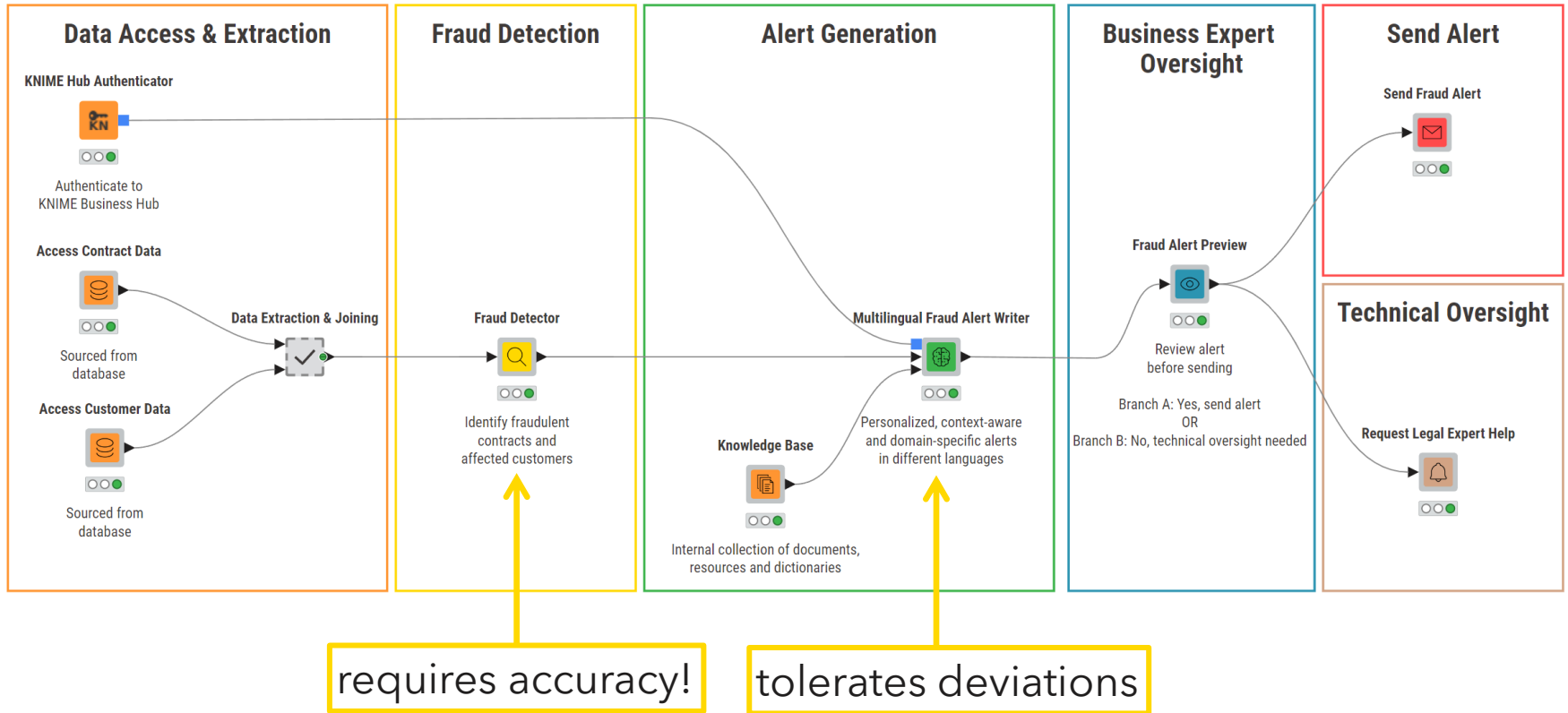
Agent Prompter



OpenAI Functions Agent Creator

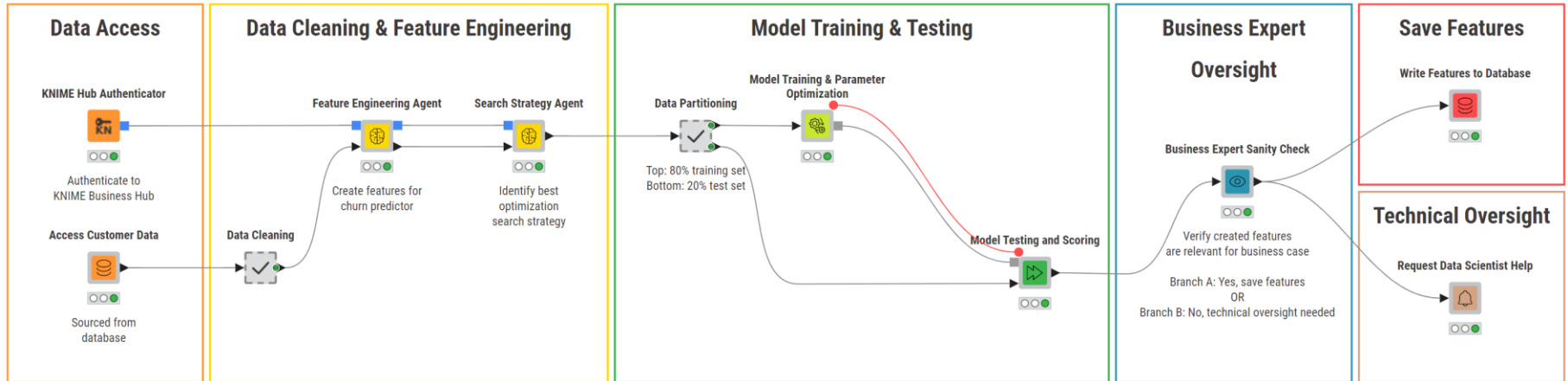


Augment Analytics: Fraud Detection & Alerting

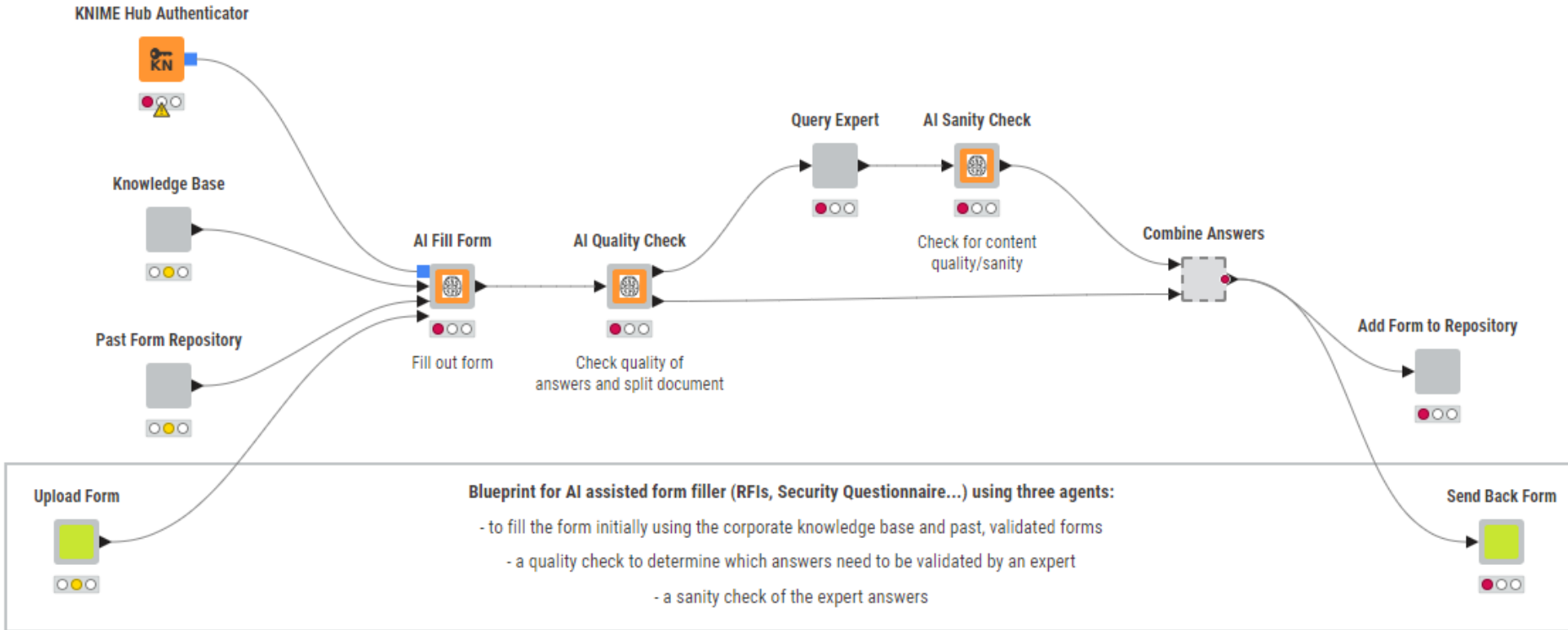


Augment Analytics with Data Science Expertise

Agentic AI for Feature Engineering of a Churn Prediction ML model



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



Elevate

Get Help Creating Workflows

Amplify

Augment Analytics

Delegate

Automate Mundane Tasks

Safeguard

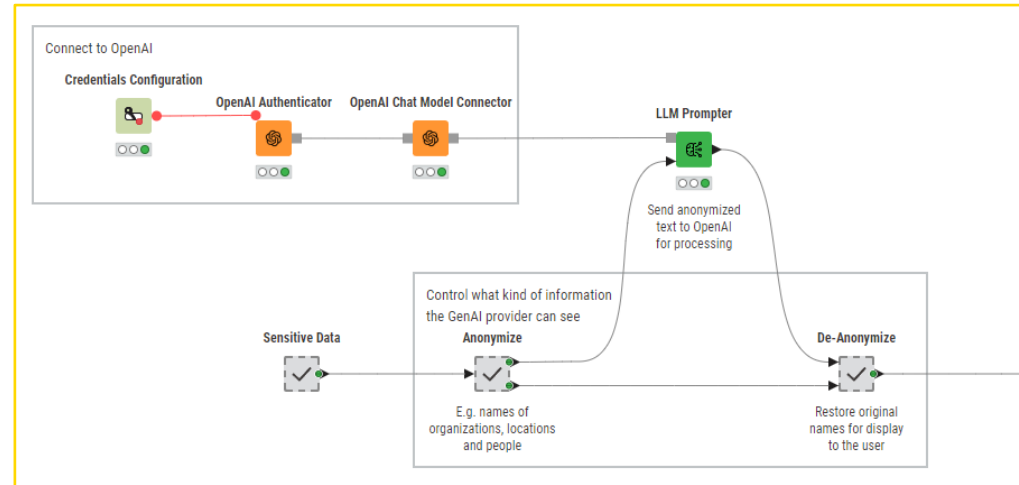
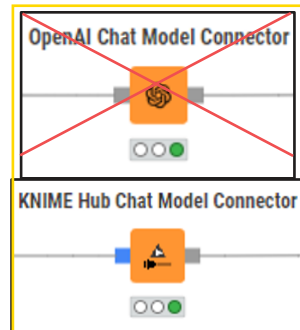
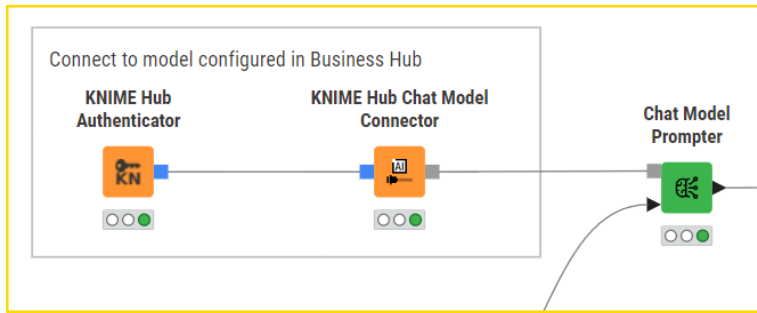
Guardrail AI

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



Elevate

Get Help Creating Workflows

Amplify

Augment Analytics

Delegate

Automate Mundane Tasks

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

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.

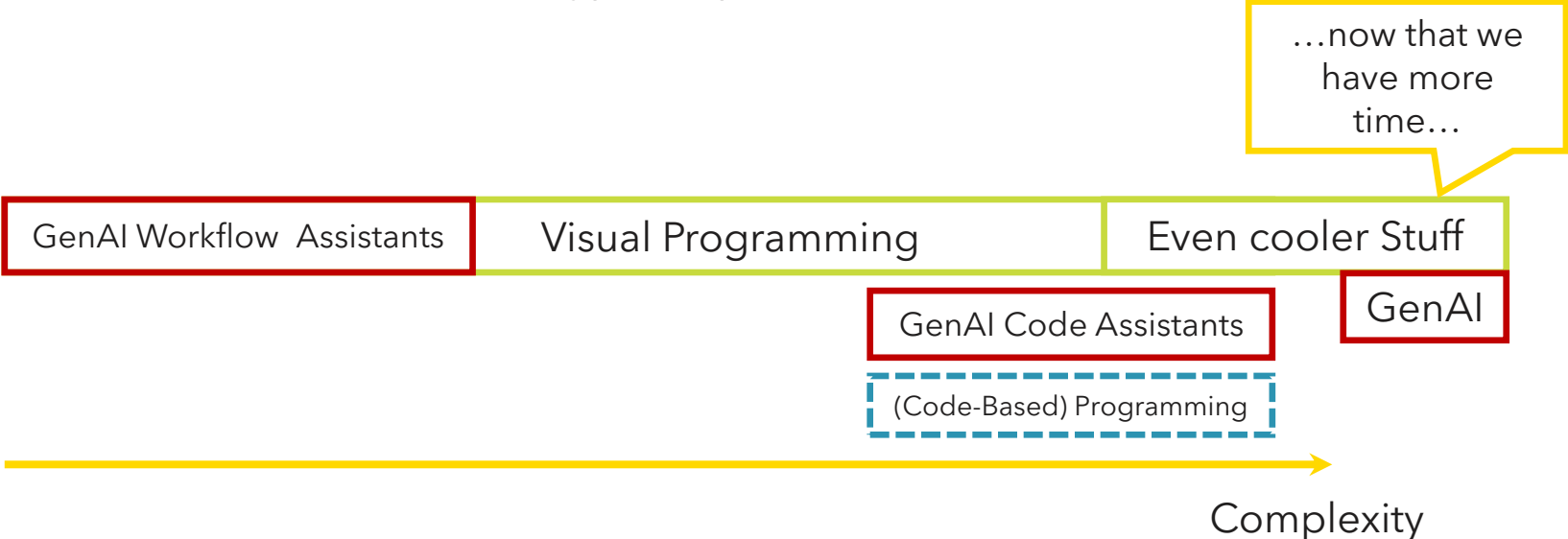
Lightweight model
Specifies the model used for simpler tasks like K-AI's Q&A mode.

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!