Proprietary + Confidential

Agenda

01 Al Applied by NextNovate

02 Securing AI - A CISO's Perspective by Qodea

03 Building Conversational Agents with Google Cloud by Xebia

Proprietary + Confidentia

01 Al Applied Build vs Use





Google Cloud

About us

Founded

2016

Great

Place

Work.

Certified

APR 2024-APR 2025

То

The Human Side of Cloud

Customer

114 +

2020 / 2021 / 2022 / 2023

For NextNovate, the cloud is all about **People**. Whether we're implementing Google Cloud Platform or Google Workspace, our focus is on empowering employees to work smarter.

Support

82.000+

users

Premier Partner

Google Cloud

Specialization

Work Transformation - Enterprise

Employees

30 +

NextNovate was founded in 2016 and has been around for approximately 8 years. In these years, NextNovate has grown from an ambitious start-up to a serious player in the field of Google Cloud Platform and Workspace.

 We co-design digital work environments that people want to continue using them at home.

And they can!

 Our mission is to empower our customers by enhancing their digital capabilities, fostering employee engagement, boosting productivity and enabling employee -led innovation.



About me

Stefan Hogendoorn

Chief Geek

Founded	Customer	Years in IT	In ecosystem since
1972	1000+	30+	2008

LinkedIn: in/stefanhogendoorn

Email: stefan@nextnovate.com

Pitfalls of AI projects

Let's build an AI solution quickly!

Can't we get GenAl to do this? Let's train our own model

(even better when talking about LLMs)

How to avoid these pitfalls?

Easy: Sit through this session and pay attention!



Understanding the AI landscape

Different requirements for different personas. One tool does not fit all.

Business Users

Easy tools to support common tasks. Often predefined or off the shelve. ð

Data Engineers

Complex data processing, applied AI and model deployment and general monitoring requirements. Data Scientists

Complex problem solving, specialised tools and advanced monitoring requirements.



Tool category	Business user	Data Engineer	Al Engineer
No-code/low code	Easy to start with and to quickly build something	Limited usage, exploration purposes mainly	Mainly for prototyping
Vertex Al AutoML Solutions	Dev skills needed (or a good Al helper)	automating model selection or hyperparameter tuning	establish baseline models quickly,handle easy parts of the ML workflow.
ML Frameworks	To complex for a non- technical business user	May utilize frameworks indirectly through tools that abstract away complexities.	Can leverage frameworks like PyTorch and TensorFlow Vertex AI supports these integrated ML frameworks
Data Engineering	To complex for a non- technical business user	Data preprocessing, transformation, and pipeline management. Vertex Al integration with BigQuery.	Data acquisition, preparation, and feature engineering. Vertex AI provides capabilities to use Cloud Storage or NFS share for custom training
MLOps	Running models is generally out of scope	Can leverage MLOps for monitoring model performance or understanding model insights, often through dashboards.	Heavily reliant on MLOps platform for managing models, orchestrating workflows and performance. Vertex AI is Google's MLOps platform.

The build vs Use dilema

Build

Developing custom ML models using frameworks like TensorFlow or PyTorch, tailored to your specific needs

Use

Utilizing pre-built models, APIs, or platforms like Vertex AI that offer readyto-use solutions. Consider the complexity of the task with regards to data, availability of knowledge, budget and time.

Also consider the maintainability and scalability of your custom solution and how to control costs.



Build

Use

1. Business Objectives	Aligns perfectly with specific and potentially unique requirements.	May not fully address highly specialized or niche business needs.
2. Data & Task Complexity	Suitable for complex tasks and data with unique characteristics	Well-suited for simpler tasks and structured data,
3. Expertise & Resources	Requires significant ML expertise, development resources, and infrastructure.	Can be more efficient with limited resources.
4. Time to Deployment	Typically involves a longer development cycle.	Offers faster time-to-market.
5. Control & Customization	Provides complete control over the model architecture, training process, and deployment environment.	May offer limited flexibility.
6. Maintenance & Scalability	Requires ongoing maintenance and updates.	Often handles maintenance and updates automatically.

Leverage the power of existing solutions

Business value		Development Effort		Production	
Focus on Core Business Value	Lowered Entry Barrier to ML	Access to State-of-the- Art Technology	Cost- Effectiveness	Simplified Maintenance & Scalability	Reduced Development Time & Effort
		Pre-trained models and services are often built on cutting-edge research and technology, providing access to high- performing solutions.	Utilizing existing solutions prove more cost-effective compared to building custom models from scratch, especially for prototyping or common ML tasks.	Existing solutions often come with built-in features for model management, monitoring, and scaling, reducing the operational overhead.	Existing solutions significantly accelerate the ML development lifecycle, enabling faster time-to-market and reducing the need for extensive coding and infrastructure setup.





Examples of existing solutions







Pick the right tool

"Not invented here" is not such a bad thing!

Key Considerations for Tool Selection:

Project Requirements: Clearly define the project's goals, the nature of the data, the desired model performance, and any specific constraints.

Technical Expertise: Assess the team's ML expertise and familiarity with different frameworks and tools.

Resource Availability: Consider the available computational resources, budget, and time allocated for the project.

Scalability and Maintainability: Factor in the long-term requirements for model deployment, monitoring, and maintenance.



Wrapping things up

Make sure you have answers to the following questions





Thank you for your attendance

Stefan@nextnovate.com

in/stefanhogendoorn

02 Securing Al - A CISO's Perspective.







Securing AI - A CISO's Perspective

Mike Smith

Director of Engineering & Security



So, what is the Challenge?







Google Cloud

Partner

How can they help?



Infrastructure Guardrails & AI Enhanced Security Operations







Conclusion - Always be proactive with Security, it pays dividends!



- <u>https://owasp.org/www-project-ai-security-</u> and-privacy-guide/
- <u>https://owasp.org/www-project-top-10-for-large-language-model-applications/</u>



- <u>A Sensible Regulatory Framework for Al</u> <u>Security</u>
- <u>https://atlas.mitre.org/resources/ai-security-</u> 101

General

- <u>https://www.nist.gov/itl/ai-risk-management-</u> <u>framework</u>
- Guidelines for Secure Al System Development, UK

National Cyber Security Centre







03 Building Conversational Agents with **Google Cloud**



Kebia

Building Conversational Agents with Google Cloud

Sander van Donkelaar Jetze Schuurmans





Jetze Schuurmans

Machine Learning Engineer Xebia Data



Sander van Donkelaar

Machine Learning Engineer Xebia Data



"The integration of the web into business processes will be gradual, but its ultimate impact on profitability and efficiency will be profound" Howard Schultz. 1997

"The early promise of the internet is marred by a reality where many businesses struggle with its complexities and fail to realize anticipated efficiencies" **Clayton Christensen**, **1997**

systemic impact because i to do ever Ma, 2017

"The AI and robots are going to kill a lot of jobs because in the future, these machines will be able to do everything better than human beings" **Jack**



"The real danger is not that AI will destroy us, but that companies will fail to adapt and harness AI effectively, leading to lost opportunities & competitive disadvantage" **Dario Gil, 2020** Fast AI adoption positions your company at the forefront of capturing most benefits

Conversational Agents

Software that interacts with users via natural language

Conversational Agents are like virtual human contactcenter agents: they handle concurrent conversations with your end-users and can perform specific actions



Customer Support



Automation



Employee Assistance

Types of Conversational Agents



What are the common challenges?

Building LLM applications in production is not easy

			Analytics
	Automation		Dashboarding
Networking	CI/CD		
VPC		Memory	Versioning
	Document Retrieval	NaSOL databasa	
	Vector Database		Prompt / Model Registry
Blob / object Storage		Development Suite	LLM Application
SQL Database	Model Training	Prompt Engineering	Orchestration
	Embedding Model	Performance Evaluation	
	Completions Medal		Grounding
Document Processing	Completions Model	App Development	Retrieval
Document Parser	Model Serving	Monitoring	Ranking
Chunking Mechanism	Model Endpoints	Cloud Logging	Tools

Most companies are not getting the benefits of generative AI



Customer Engagement Suite

Fully managed, unified platform for building AI-powered CX operations



Customer Reference: ING

Improve customer experience by instant answers to questions related to daily banking

Shorter time to market: Generative AI simplifies chatbot development compared to traditional, rule-based agents.

Better performance: the conversational agent provided improved deflection and customer satisfaction rates.

Now (pilot) Chatbot handles around 600 conversations per day in NL. Next (scale) Handle over 5000 conversations per day. Scale across multiple countries.



1. The core of the agent, combining generative and traditional features



1. The core of the agent, combining generative and traditional features



1. The core of the agent, combining generative and traditional features



Best of both worlds: combine precise conversation controls with generative features

Flows: Flows consist of conversational paths / journeys



Best of both worlds: combine precise conversation controls with generative features

Flows: Flows consist of conversational paths / journeys

Pages: states are represented by pages: each page represents a "step" in the conversational journey.



Best of both worlds: combine precise conversation controls with generative features

Flows: Flows consist of conversational paths / journeys

Pages: states are represented by pages: each page represents a "step" in the conversational journey.

Routing: Detect intents or use LLMs to route the query towards the appropriate flow or page.



Best of both worlds: combine precise conversation controls with generative features

Flows: Flows consist of conversational paths / journeys

Pages: states are represented by pages: each page represents a "step" in the conversational journey.

Routing: Detect intents or use LLMs to route the query towards the appropriate flow or page.

Responses: can be fixed or Al-generated.



2. Use Generators for generative AI

Combine Gemini with your custom prompts to generate responses at runtime

Routing: use the output of generators to determine the next step

Response Generation: use generators to create dynamic LLM based responses

Data Extraction: extract or edit structured JSON objects from conversations.

Security: use generators as guardrails

Generators allow you to use generative AI models to generate dynamic responses or text that can be used during fulfillment.

- Display name*

ConversationSummarization

Configure the text prompt that is sent to the generative model. Mark words as placeholders by prepending \$. Placeholders are associated with session parameters in fulfillment and replaced by session parameter values during execution. You can also use the built-in placeholders \$conversation and \$last-user-utterance.

Text prompt* You are an expert at summarizing conversations between a User and an Agent. When providing the summary, always start with "Dear joe@example.com conversation summary is as follows:" Provide a summary in a few bullet points. Try to be as brief as possible with each bullet point, and the low point of the conversation

Model configuration

2. Generators as guardrails

Call Gemini from your conversational agent to prevent harmful responses



3. Data Stores: generate responses based on your own data

An end-to-end solution to automatically index- and retrieve data



3. Finetune the responses of your agent

The UI enables non-technical users to develop prompts

Data store prompt

Providing this information can improve the quality of answers generated from data store content and make them feel more your brand. Learn more

Provide text in English even if your agent is configured in another language.

Agent name	Example:
Agent identity	Your name is the ACME Virtual Assistant , and you are a helpful and polite AI Assistant at ACME Co, a fictional e-commerce site. Your task is to assist humans on the company
Company name*	website.
Annes	Your prompt:
Company description	helpful and polite customer assistant at Xebia Airlines, xebia-airlines.com. Your task is to
Agent scope You will be used at the company website to assist clients in handling user queri	assist humans You will be used at the company website to assist clients in handling user queries.

The will be doed of the comband menore of policy or drawing and does

user queries

4. Integrate it with your cloud- and data infrastructure

Extensive Support of Data Sources

Integrate your existing data layer as storage backend for your data stores.

Simplified Export

Conversations can are exported to BigQuery and Cloud Logging with one click.

Alerting & Observability

Create alerts based on logs, or route logs to third party systems.

Webhooks

Let your agent call APIs such that it can perform specific actions

ative sources					
(•	E38	\$	RPI	9
Website Content Automatically crawl public website content from a list of URL patterns you define.	BigQuery Import data from your BigQuery table.	Cloud Storage Import data from your storage bucket.	Healthcare API (FHIR) Import FHIR store data from your Cloud Healthcare API dataset. This allows you to create an app on your clinical data.	API Import data manually by calling the API.	Cloud SQL Import data from your Cloud SC table.
SELECT	SELECT	SELECT	SELECT	SEE DOCUMENTATION 2	SELECT
*		\$			
Spanner PREVIEW	Bigtable PREVIEW	Firestore	AlloyDB PREVIEW		
Import data from your Spanner table.	Import data from your Bigtable table.	Import data from your Firestore collection.	Import data from your AlloyDB cluster.		
SELECT	SEE DOCUMENTATION	SELECT	SELECT		
orkspace sources					
4	M	B	1	#	
Google Drive	Google Gmail PREVIEW	Google Sites PREVIEW	Google Calendar PREVIEW	Google Groups PREVIEW	
Link to your organization's drive	Link to your organization's Gmail	Link to your organization's Sites	Link to your organization's Calendar	Link to your organization's Groups	

4. Evaluating your agent

Collect feedback and log conversational interactions to enable advanced analytics and monitoring



Combine domain-expertise with technical skills



Low-code

Non-technical users can use the console to develop conversational agents



Code-first

Technical users can leverage the client SDKs, or use IaC tooling such as terraform to build agents.

Low-code without sacrificing best practices





CI/CD

Every component can be provisioned using Terraform Client SDKs and REST APIs can be used to fully automate the deployment process



Version Control

Agents are fully integrated with Github and can be exported and restored from .json files

Unlock GenAl with Conversational Agents on 🙆 Google Cloud

Best of Both Worlds

Combine precise conversation controls with generative features.

Low Code

Make use of the potential of your entire workforce by combining domain-expertise with technical skills.

Cloud Native

Reap the full benefits of the cloud. Easily integrated with existing data- and cloud applications.





Short time to value

Low Costs of Ownership

Questions?



Thank You