Skill Coach Hackathon

Guidelines for Innovation & Responsibility

A collaboration between Škoda Auto a.s. and 42 Prague

Welcome, Innovators!

We're thrilled to have you!

- **Our Goal:** To foster creativity, collaboration, and rapid innovation.
- **Your Challenge:** Gain new insights from our employee skills and job description data.
- **Your Mission:** Build amazing, viable solutions that could shape our future.

These guidelines are here to empower you while keeping our data secure.

Let's build the future, responsibly.

🔒

The Core Principle: Data Security

This is our #1 Rule.

You are working with anonymized but sensitive company data.

Protecting it is your **top priority**.

The Golden Rule
The provided dataset and any derivatives of it must not leave the designated hackathon environment.

🛡️🧠

Your Toolkit: Secure AI Endpoints

Leverage LLMs Safely

We have provisioned secure, internal LLM endpoints for you.

- **Models:** Access powerful models like **GPT-4o** and others.
- **Infrastructure:** Hosted securely on our internal Azure environment.
- **How it works:** A single API endpoint that is fully compatible with the standard OpenAI API.

Your welcome packet contains your unique API URL and authentication credentials.

✅

The *Only* Approved Method

Using LLMs on Hackathon Data

Interacting with our provided **secure, internal API endpoint** is the **ONLY** approved method for using a Large Language Model on the hackathon dataset.

- Analyze data
- Summarize text
- Generate insights

...all within our secure, sandboxed environment.

Recommended UI: OpenWebUI

Quick Start: OpenWebUI

A self-hosted, open-source chat interface you can point to our secure endpoint.

GitHub: github.com/open-webui/open-webui

Recommended Install (Docker):

```
docker run -d -p 3000:8080 \
  -e OLLAMA_BASE_URL=YOUR_API_ENDPOINT_URL \
  -v open-webui:/app/backend/data \
  --name open-webui --restart always \
  ghcr.io/open-webui/open-webui:main
```

*(**Note:** The `OLLAMA_BASE_URL` variable works for any OpenAI-compatible API, including ours!)*

Access it at: **http://localhost:3000**

Recommended UI: LibreChat

Alternative Option: LibreChat

Another powerful, self-hosted UI for interacting with our LLM endpoint.

GitHub: github.com/danny-avila/LibreChat

Recommended Install (Docker Compose):

1. Create a `docker-compose.yml` file from their official setup guide.
2. Create a `.env` file and configure your custom endpoint.
3. Run:

```
docker compose up -d
```

(Local install without Docker is complex and recommended for advanced users only.)

Critical Restriction: External AI Tools

Using Public AI (ChatGPT, Copilot, etc.)

We encourage you to use external tools for general tasks, but with one critical rule:

You must NEVER send, paste, or upload the provided sensitive data to any external, public, or third-party service.

Violation will result in immediate disqualification.

How to Use External AI Safely

The Key: Abstraction & Anonymization

❌ **BAD (FORBIDDEN)**	✅ **GOOD (ALLOWED)**
Pasting into public ChatGPT:	Asking public ChatGPT:
"Here is a list of 50 job descriptions from our company. Cluster them by seniority."	*"Write a Python script using scikit-learn that takes a list of text documents and clusters them using TF-IDF and K-Means."*

Use external tools for code, ideas, and assets—not for data processing.

☁️

Tech Guidelines: Cloud Infrastructure

Build for Our Ecosystem

- **Primary Cloud:** Škoda Auto is an **Azure** ecosystem. We strongly encourage using the provided Azure services.
- **Other Clouds (AWS, GCP)?** Allowed only if:

  1. **Unique Capability:** It offers a clear, justifiable advantage not available on Azure.
  2. **Platform Agnostic:** Your solution is fully **containerized** (e.g., Dockerfile) to ensure it can be deployed anywhere.

👍⚠️

Tech Guidelines: Open Source Licenses

Choose Your Dependencies Wisely

For a project to be commercially viable, it must use permissively licensed software.

✅ Highly Encouraged (Permissive)

✔ **MIT** (NumPy, Pandas, .NET)
✔ **Apache 2.0** (TensorFlow, Spark)
✔ **BSD** (PyTorch, scikit-learn)

⛔ Generally Avoid (Strong Copyleft)

❌ **GPL**
❌ **AGPL** (Affero General Public License)

Your Responsibility: Check the `LICENSE` file before adding a new dependency! When in doubt, ask an organizer.

🔒

Tech Guidelines: Version Control Hygiene

Secure Coding with Git & GitHub

Rule 1: NEVER Commit Credentials

- **Problem:** Hardcoding API keys or passwords exposes them in your repository's history forever.
- **Solution:** Use **environment variables**. Load them from a `.env` file for local development.

Rule 2: NEVER Commit the Dataset

- **Problem:** The sensitive `.csv` or `.json` data file should never be in your repository.
- **Solution:** Add the data files and your `.env` file to your `.gitignore` file.

⚠️

A Note on Automation Tools

Advanced Local Setup: Pinokio Browser

- **What it is:** A tool to simplify installing and running local AI applications from GitHub.
- ⚠️ **IMPORTANT WARNING:** Tools like Pinokio may require **root/administrator access** to your machine.
- Granting these permissions carries inherent security risks.

Use these tools at your own risk.

We recommend using isolated environments (like VMs or containers) if you choose to explore them.

Hackathon Code of Conduct

Our Etiquette

🤝 **Collaborate and Respect:** Support your teammates and respect all participants.

💡 **Innovate and Have Fun:** Think outside the box, experiment, and don't be afraid to fail.

❓ **Ask for Help:** Mentors and organizers are here for you. If you're stuck, just ask!

🏆 **Present with Pride:** Show us what you built, what you learned, and the potential of your idea.

Questions?

We are incredibly excited to see what you create.

Happy Hacking!