Skill Coach Hackathon: Guidelines for Innovation & Responsibility

A collaboration between Škoda Auto a.s. and 42 Prague
Welcome, Innovators!

Welcome to the Skill Coach Hackathon! We are thrilled to see what you will build. The goal of this event is to foster creativity, collaboration, and rapid innovation by tackling a real-world challenge: gaining new insights from our employee skills and job description data.

To ensure this event is both productive and secure, we've put together these guidelines. Please read them carefully. They are designed to empower you to build amazing things while protecting our company's data and ensuring your projects are viable for future development.

Let's build the future, responsibly.

---

1. Core Principle: Data Security & Confidentiality

This is our most important rule. You will be working with a dataset containing anonymized but sensitive information related to employee roles, skills, and certifications. Protecting this data is your top priority.
The Golden Rule: The provided dataset and any derivatives of it **must not leave the designated hackathon environment.**

1.1. Provided Secure AI Endpoints & Recommended Tools

To enable you to leverage the power of Large Language Models (LLMs) safely, we have provisioned secure, internal LLM endpoints giving you access to powerful models like **GPT-4o** and others hosted on our Azure infrastructure.

- **What they are:** These are company-approved, sandboxed instances of powerful language models. They are exposed via a single API endpoint that is fully compatible with the OpenAI API standard.
- **How to use them:** You can interact with this endpoint via an API to perform analysis, summarization, and other tasks directly on the sensitive dataset. Your hackathon welcome packet will include the specific API URL and your authentication credentials.
- **This is the ONLY approved method for using an LLM on the provided hackathon data.**
Recommended UIs for Chatting with the Endpoints:

To get started quickly, you can use a self-hosted, open-source chat interface. These UIs can be pointed to our secure LLM endpoint. We recommend the following options:

- **OpenWebUI**
  - **GitHub:** https://github.com/open-webui/open-webui
  - **Recommended Install (Docker):** This is the fastest way to get started. To connect OpenWebUI to our custom endpoint, you'll set the `OLLAMA_BASE_URL` environment variable. **Note:** Despite its name, this variable is used by OpenWebUI to connect to *any* OpenAI-compatible API, including our Azure-hosted one.

```bash
# Replace YOUR_API_ENDPOINT_URL with the URL we provide
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
```

You can then access the interface at `http://localhost:3000`.

  - **Local/Pip Install (Alternative):** For those who prefer not to use Docker, OpenWebUI can be installed via pip. Ensure you have Python 3.11 or newer.

```bash
# 1. Install the package
pip install open-webui

# 2. Run the server, pointing it to our endpoint (the variable name is
```

```
OLLAMA_BASE_URL, but it works for any OpenAI-compatible API)
OLLAMA_BASE_URL=YOUR_API_ENDPOINT_URL open-webui serve
```

(On Windows, you may need to set the environment variable separately before running the `serve` command).

- **LibreChat**
  - **GitHub:** https://github.com/danny-avila/LibreChat
  - **Recommended Install (Docker):** First, create a `docker-compose.yml` file with the contents from their official setup guide. Then, create a `.env` file in the same directory and define your custom endpoint configuration as described in their documentation. Finally, run:

```bash
docker compose up -d
```

  - **Local Install (Advanced):** A local installation without Docker is possible but is significantly more complex as it requires setting up the full-stack environment (Node.js, MongoDB, etc.) manually. This is recommended only for participants comfortable with this process. For instructions, please follow the official Local Development Guide in the LibreChat documentation.

1.2. Using External AI and "AI as a Service" Tools

We encourage you to use the vast ecosystem of AI tools available today (e.g., public ChatGPT, GitHub Copilot, Claude, Midjourney, etc.), but with one critical restriction: **You must NEVER send, paste, or upload the provided sensitive data to any external, public, or third-party service.**

Doing so is a direct violation of our data security policy and **will result in immediate disqualification** of your team.
How to use external AI tools safely:

The key is **abstraction and anonymization**. You can rephrase your problem or use the tools for general-purpose tasks.

- **BAD EXAMPLE (Forbidden):**

Pasting into ChatGPT: *"Here is a list of 50 job descriptions from our company. Cluster them by seniority and required cloud certifications."*

- **GOOD EXAMPLE (Allowed):**

  Asking ChatGPT: *"Write a Python script using scikit-learn that takes a list of text documents and clusters them using TF-IDF and K-Means. The script should output the top 5 keywords for each cluster."*

  You can use external tools to generate code, brainstorm approaches, debug algorithms, or create presentation assets, but the sensitive data itself must remain within our secure environment.

---

2. Technology & Architecture Guidelines

Our goal is to create solutions that could one day be integrated into our company's ecosystem. Following these technical guidelines will increase the long-term viability of your project.

2.1. Cloud Infrastructure

Škoda Auto is primarily an **Azure** ecosystem. We strongly encourage you to use the Azure services provided for the hackathon.

However, we are open to other cloud vendors (AWS, GCP, etc.) under two conditions:

1. **Unique Capability:** The vendor offers a specific service or feature that is uniquely suited to your solution and provides a clear advantage that isn't easily replicated on Azure. You should be prepared to justify this choice in your final presentation.
2. **Platform Agnostic:** Your application is built to be completely cloud-agnostic. The best way to demonstrate this is by containerizing your solution (e.g., using a **Dockerfile** and **Docker Compose**). This ensures it can be deployed anywhere, including our Azure environment, with minimal effort.

2.2. Open Source Software (OSS) Licensing

We love Open Source! However, for a project to be commercially viable, it must be built with appropriately licensed software. Please prioritize using OSS with permissive

licenses.

Here's a quick guide to help you choose:
✅ **Highly Encouraged: Permissive Licenses**

- **MIT License:** (e.g., NumPy, Pandas, .NET Core)
- **Apache License 2.0:** (e.g., TensorFlow, Spark, Kubernetes)
- **BSD Licenses (2-Clause and 3-Clause):** (e.g., PyTorch, scikit-learn, Go)
  ⛔ **Generally Avoid: Strong Copyleft Licenses**
- **GPL (General Public License)**
- **AGPL (Affero General Public License):** (e.g., Mastodon, Ghostscript)
  Your Responsibility: Before adding a new library or dependency, check its `LICENSE` file. When in doubt, ask a hackathon organizer!

2.3. Version Control & Code Hygiene

You are encouraged to use Git and GitHub to collaborate. However, you must follow secure coding practices to protect our data and your credentials.

- **NEVER Commit Credentials:** Never hardcode API keys, passwords, or other secrets directly in your source code. Once committed, they can be exposed forever in the repository's history.
  - **Solution:** Use environment variables. A common pattern is to use a `.env` file for local development.
  - **Example (Python):** A great library for this is `python-dotenv`.

1. Install it: `pip install python-dotenv`
2. Create a file named `.env` in your project root with your secret:
   `HACKATHON_API_KEY="your_secret_key_here"`
3. Load it in your code:

<> python

```
import os
from dotenv import load_dotenv

# Load variables from .env file into environment
```

```
load_dotenv()

api_key = os.getenv("HACKATHON_API_KEY")
```

- **Crucially, you must add `.env` to your `.gitignore` file** to prevent your secrets from ever being committed.
- **NEVER Commit the Dataset:** The provided sensitive dataset (`.csv` , `.json` , etc.) must never be committed to your Git repository.
- **Solution:** Use a `.gitignore` file. Create a file named `.gitignore` in the root of your project and add the names of files or folders you wish to exclude.
- **Example `.gitignore`:**

```
<>

# Ignore data files
*.csv
*.json
data/

# Ignore environment files
.env

# Ignore Python virtual environments
venv/
```

---

3. A Note on Automation Tools

For those comfortable with local machine setup, tools exist to simplify the installation of AI applications.

- **Pinokio Browser:** Pinokio is a browser that lets you locally install, run, and automate AI applications from GitHub. Everything you can run in your command line can be automated with a Pinokio script, with a user-friendly UI.
  ⚠️ **IMPORTANT WARNING:** Be aware that tools like Pinokio may require **root/administrator access** to your machine to install packages and manage environments. Granting such permissions carries inherent security risks. **Use these tools at your own risk.** We recommend using isolated environments (like virtual machines or containers) if you choose to explore them.

---

4. Code of Conduct & Hackathon Etiquette

- **Collaborate and Respect:** This is a team event. Support your teammates, listen to their ideas, and work together. Extend that respect to all other teams and the event organizers.
- **Innovate and Have Fun:** Think outside the box! This is your chance to experiment with new technologies and bold ideas. Don't be afraid to fail and learn. The primary goal is to learn and have fun.
- **Ask for Help:** Mentors and organizers are here to help you. If you're stuck on a technical problem, have a question about the guidelines, or need any other assistance, please don't hesitate to reach out.
- **Present with Pride:** At the end of the hackathon, you'll present a short demo of your work. Focus on what you've built, what you've learned, and the potential value of your idea.

We are incredibly excited to see what you create. Let's make this a fantastic event for everyone involved.

Happy Hacking