Technical Take-Home Exercise — AI Solutions Engineer

Timebox: ~4 focused hours

1 . Objective

Design and code a **Python (≥ 3.8)** pipeline that turns a company's public web presence into a **600 – 1000-word Markdown factsheet**.

- Treat this as a mini-solutioning + coding + system-design challenge.
- Assume the factsheet will be read by Sales reps before discovery calls.
- We expect you to decide **which data points belong in the factsheet and why**; justify your choices in a short note.

2 . Data You'll Receive

A list of 6 sample companies -- all small- to mid-sized firms.

You only need to generate one factsheet to prove end-to-end functionality; the full list is provided so you can test broadly if you wish.

URL	Industry
https://dreeshomes.com/	Construction
https://www.good2grow.com/	Retail
https://silkroadmed.com/	Manufacturing & Production
https://nationalcareadvisors.com/	Healthcare
https://www.drinktractor.com/	Manufacturing & Production
https://www.darkhorse.cpa/	Financial Services & Insurance

3 . Your Task

1. Accept a single URL (from the CSV).
2. Gather open-web information — scraping pages, using free search APIs, etc.
3. Synthesize the information into a clear, narratively coherent mini-report. There are no fixed structure requirements, you are free to decide that.
4. Output one Markdown file (<company-name>.md, 600 – 1000 words).

4 . Requirements & Constraints

Area	What We Expect
Language	Python 3.8 + only.
Budget	You'll receive an OpenAI key (valid 5 days, $50) that we monitor. Use any model; manage cost yourself.
Libraries & APIs	Free/open-source packages and free-tier APIs only. No paid/proprietary services beyond the supplied key.
README (root)	*Must* include: • Dependency install (pip install … or requirements.txt) • How to set OPENAI_API_KEY • Exact command(s) to run your pipeline • 2-3 paragraphs on your design choices, success definition, and trade-offs.
Anchors	Assume pages are in English. Each company has a homepage and an "About" page.

5 . Deliverables (Public Github Repo)

A link to a public GitHub (or any VCS) repo that we can access.

```
your-solution
├─ README.md                ← instructions & design notes
├─ requirements.txt         ← or pyproject.toml
├─ src/                     ← your code (any structure)
├─ companies.csv            ← provided input list
└─ <company>.md             ← one sample factsheet (600 – 1000 words)
```

6 . What We're Looking For

Reviewers will broadly consider:

- Sound reasoning for chosen data points & success criteria.
- Code clarity, project scaffolding, and reproducibility.
- Effective, optimal use of specific LLMs.
- Thoughtful handling of ambiguous or missing data.

7 . Time & Resource Expectations

Plan to spend **no more than ~4 hours**. Prioritize a working, extensible prototype over elaborate infrastructure.

You are allowed to use any AI-assisted coding tools to build this. You can use the provided API key for that.

You are also encouraged to find a small, permissively licensed public repo and build atop it. You need to provide attribution so that we can evaluate how your changes and solution design address the problem at hand.

8 . Submission

Email the link to your repository holding the solution within the agreed timeline. Include any clarifying assumptions in your README.

Good luck and have fun!