

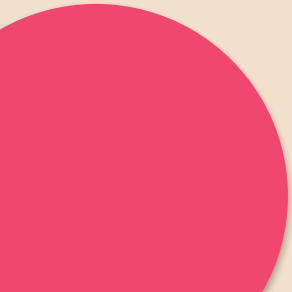


# Les 18

## Advanced AI Toolbox

Voice, vision, image, local, edge

Wat er nog meer mogelijk is — de laatste les



Terugblik

# 18 lessen samen

Lessen 1-10

Foundations: Next.js, TS, Supabase basics

Les 11

AI SDK basics

Les 12

Tool Calling

Les 13

Cursor + Vercel deploy

Les 14

Agents

Les 15

RAG + Embeddings

Les 16

MCP servers

Les 17

Production Polish

Les 18

**Vandaag: Advanced AI Toolbox**

Vandaag: alles wat we nog NIET hebben gezien. Demo-driven, kort.

## Planning

# Vandaag — 180 minuten

|                                              |        |
|----------------------------------------------|--------|
| Terugblik 18 lessen                          | 10 min |
| Theorie: AI landschap 2026                   | 15 min |
| <b>LIVE DEMO 1 — Voice (Whisper + TTS)</b>   | 25 min |
| <b>LIVE DEMO 2 — Vision (GPT-4o)</b>         | 20 min |
| <b>Pauze</b>                                 | 15 min |
| <b>LIVE DEMO 3 — Image generation (Flux)</b> | 25 min |
| <b>LIVE DEMO 4 — Local LLMs (Ollama)</b>     | 25 min |
| <b>LIVE DEMO 5 — Edge AI + AI Gateway</b>    | 20 min |
| Resources + community                        | 10 min |
| Vragen + afsluiting cyclus                   | 15 min |

## Theorie

# Het AI-landschap 2026

### Modellen per provider:

|                           |                                              |
|---------------------------|----------------------------------------------|
| OpenAI                    | GPT-5, o3, GPT-realtime, DALL-E 4, Whisper   |
| Anthropic                 | Claude Opus/Sonnet/Haiku 4.5+                |
| Google                    | Gemini 2.5 Pro (2M context), Flash, Imagen 3 |
| Meta                      | Llama 4 (open-source) + Llama 4 Vision       |
| xAI                       | Grok 4, Grok 4 Vision                        |
| Black Forest / ElevenLabs | Flux 1.1 Pro (image), ElevenLabs (TTS)       |

### Modaliteiten:

|      |        |       |           |       |
|------|--------|-------|-----------|-------|
| text | vision | voice | image gen | video |
|------|--------|-------|-----------|-------|

### Modi om te draaien:

- Cloud API      direct via provider
- Edge      Vercel AI Gateway, Cloudflare
- Lokaal      Ollama, llama.cpp, LM Studio

## Theorie

# Waarom multimodal — tekst is niet genoeg

### Voice input

Handsfree, sneller dan typen, mobile-dominant

### Vision

User upload foto — AI analyseert (OCR, defect, beschrijving)

### Image gen

Assets ter plaatse — landing pages, social, avatars

### Voice output

Voor accessibility, cars, IoT

### Real-time

Conversational, geen wachten op response

**In je apps: denk niet 'chat only'**

Vraag je af welke modaliteit het best past per feature.

## LIVE DEMO 1

# Voice — Whisper + TTS — ~25 min

Transcriptie (audio -> text):

```
import { experimental_transcribe as transcribe } from "ai";
const result = await transcribe({
  model: openai.transcription("whisper-1"),
  audio: audioBlob,
});
```

TTS (text -> audio):

```
const audio = await speak({
  model: openai.speech("tts-1"),
  text: "Welkom bij Polderfest!",
  voice: "alloy",
});
```

### Demo-app: voice-chat met Polderfest

Mic-button -> spreek -> Whisper -> LLM -> TTS -> spreekt antwoord.

Bonus: Realtime API voor lagere latency + interruptions (WebRTC).

## LIVE DEMO 2

# Vision — GPT-4o foto-analyse — ~20 min

```
const result = await generateText({
  model: openai("gpt-4o"),
  messages: [{
    role: "user",
    content: [
      { type: "text", text: "Beschrijf wat er op deze foto staat." },
      { type: "image", image: imageUrl }, // of base64 buffer
    ],
  }],
});
```

### Demo-cases:

- Foto eten -> ingrediënten
- Screenshot error -> uitleg
- Handtekening -> stijl

### Productie use cases:

- Defect-detectie fabriek
- OCR formulieren
- Visual search
- Alt-tekst (a11y)

**Cost: vision is duurder dan text — check pricing per provider.**



# **Pauze**

15 minuten

## LIVE DEMO 3

# Image generation — Flux — ~25 min

```
import { experimental_generateImage as generateImage } from "ai";
import { fal } from "@ai-sdk/fal";

const { image } = await generateImage({
  model: fal.image("fal-ai/flux/dev"),
  prompt: "Poster voor Polderfest 2027, retro stijl, oranje + cream",
  size: "1024x1024",
});
```

### Use cases:

- Landing page hero
- Social media posts
- Avatar per user
- Product mockups

### Tips:

- Specifieke prompts werken beter
- Flux Pro = kwaliteit
- Flux Schnell = snel
- \$0.02-0.10 per image

Alternatieven: DALL-E 4 (OpenAI), Imagen 3 (Google), Midjourney via API.

## Local LLMs — Ollama — ~25 min

```
brew install ollama
ollama pull llama4:8b
ollama run llama4:8b # runt op je laptop, geen internet
```

```
import { createOllama } from "ollama-ai-provider";
const ollama = createOllama();
const result = await generateText({
  model: ollama("llama4:8b"),
  prompt: "...",
});
```

### Wanneer lokaal:

- Privacy — geen data extern
- Cost — geen API-bill
- Offline werkt
- Snel op M-series Mac

### Nadelen:

- Kwaliteit < GPT-5 voor complex
- Hardware vereist (8GB+ RAM)
- Trager voor grote modellen
- Geen multimodal voor klein

## LIVE DEMO 5

# Edge AI + Vercel AI Gateway — ~20 min

Vercel AI Gateway — één endpoint, alle providers:

```
import { vercel } from "@ai-sdk/vercel";

const result = await generateText({
  model: vercel.fallback([
    "openai/gpt-4o",
    "anthropic/claude-sonnet-4.5", // fallback bij outage
  ]),
  prompt: "...",
});
```

Cloudflare Workers AI — modellen op edge, ms latency wereldwijd:

```
const result = await generateText({
  model: workersai("@cf/meta/llama-3.1-8b-instruct"),
});
```

### Wanneer edge:

- Globale app — latency wereldwijd belangrijk
- Cost-sensitive — vrijgeveige free tiers
- Provider-agnostic — geen vendor lock-in

Wat nu?

# Resources + community — hoe blijf je leren

## Communities:

- r/LocalLLaMA — Reddit
- AI Engineer Foundation Discord
- Latent Space community
- Hugging Face Discord
- AI SDK Discord (Vercel)

## Nieuwsbronnen:

- Simon Willison's blog (best AI-blog)
- Latent Space podcast (wekelijks)
- The Rundown AI (newsletter)
- Karpathy, swyx, AnthropicAI op X

## Voor diepere kennis:

- Karpathy's 'Zero to Hero' — neural nets vanaf nul (YouTube, gratis)
- Andrew Ng courses — Coursera, structureel
- Hugging Face NLP course — gratis, hands-on

**Conferenties: AI Engineer Summit (SF, jaarlijks), NeurIPS, ICLR**

Het veld beweegt snel — kies 2-3 bronnen en blijf consistent.

## Praktijk

# Eindopdracht — volgende stap

### Wat je nu hebt:

- v Volledige toolkit voor moderne AI-apps
- v 17 lesopdrachten + huiswerk afgerond
- v Eigen mini-projecten gebouwd

### Eindopdracht moet bevatten:

- Next.js + TypeScript + Tailwind
- Supabase (DB + Auth + RLS)
- Vercel AI SDK met Tool Calling
- Externe API
- Deployed naar Vercel

### Tips:

- Werk gestaag — 8-10 uur/week voor 4-6 weken
- Push klein en vaak; vraag hulp tijdig
- Begin met je voorstel (via Teams)

## Afsluiting

# Bedankt en succes

### Wat we samen gedaan hebben:

- v 18 lessen, ~54 uur klassikaal
- v Van localhost naar productie
- v Van simple fetch naar agents + RAG + MCP
- v Van mock-data naar app met observability + evals

### Wat je nu zelfstandig kunt:

- Volledige AI-app van scratch naar productie
- Multi-modal (voice, vision, image gen)
- Observability + evals + security
- Open-source AI ecosysteem navigeren

**Doe het ding waarvan je dacht: dat kan ik niet.**

Vraag hulp, deel je werk, blijf bouwen.



# Vragen?

Open ruimte — ~15 min

## Voor:

- Vragen over leerstof
  - Vragen over eindopdracht
  - Carrière in AI development
  - Feedback op de leerlijn
- 