

Wake-Word Training Research Notes

OpenWakeWord

- **Training:** Google Colab notebook available (~1 hour)
- **Data Requirements:** Minimum several thousand samples recommended
- **Negative Data:** ~30,000 hours recommended for low false-accept rate
- **Language:** Currently English only
- **Python Version:** 3.10+
- **License:** Apache-2.0
- **Stars:** 1.6k
- **Windows:** Compatible (Python-based)
- **Cost:** Free & Open Source

Porcupine (Picovoice)

- **Training:** Type phrase on Picovoice Console (seconds)
- **Method:** Transfer learning (no data collection needed)
- **Windows:** Fully compatible
- **Cost:** Free tier available
- **Accuracy:** Production-ready
- **Multi-language:** Supports German, French, Spanish, Italian, Japanese, Korean, Portuguese

Key Findings

1. Porcupine = Easiest (no recordings needed, transfer learning)

2. OpenWakeWord = Most flexible (open source, customizable)
3. Snowboy = Deprecated (use Snowman reimplementation)
4. “Computer” = Good wake word (2-4 syllables, distinctive phonemes)