Jakub Towarek

kuba@towarek.pl | +48 796 110 555 | github.com/3Qax | LinkedIn

EDUCATION

Newcastle University — Newcastle upon Tyne, UK
Computer Science (incomplete) — *2019*

WORK & LEADERSHIP EXPERIENCE

Wisent — Lublin, PL
Founding Engineer — *May 2025-present*

- Single-handedly built the native iOS app for the Wisent platform, with on-device LLM, image-generation, and TTS inference via MLX
- Build the inference-time runtime that exposes concept directions for read, write, and clamp operations
- Build and maintain the contrastive-pair concept-dataset pipelines and per-layer activation capture spanning HuggingFace transformer stacks at sizes from 1B to 405B
- Architect Wisent's multi-cloud infrastructure: European HPC allocation for training, a Python/FastAPI backend on AWS (Lambda, EventBridge, EC2, S3, CloudFront), and the app.wisent.ai web application
- Re-architected backend from a monolithic A100-coupled service into independently deployable CPU API, inference, and image services, removing model-load from routine releases and reducing deploys from 36-44 min to ∼8 min (∼78-82% faster) while eliminating routine deploy-time A100 usage (∼36-44 A100-min saved per deploy; 100% GPU deploy-cost reduction)
- Ported on-device speech/LLM inference stack to iOS by implementing NeuCodec in Swift/MLX and extending Swift Transformers/CoreML tokenization support, covering a 24kHz / 50Hz neural audio codec, 2048→1024 projection, 12-layer / 16-head Transformer, ISTFT decoding, safetensors weight loading, and Python-reference parity tests, enabling server-free real-time TTS/voice inference on Apple devices
- Led cross-platform, multi-layer safety/feed moderation across iOS, Android, and web by migrating home/search/category feeds to server-filtered RPCs, adding automated NSFW/pending gating plus an admin moderation panel with calibration tooling, and replacing duplicated client filters with consistent policy enforcement across 3 product surfaces

SumUp — Berlin, DE
iOS Engineer, Online Store Tribe — *Feb 2023-Feb 2025*

- Owned the full development cycle of key merchant-facing features, including order management used by over 100,000 merchants, within SumUp's pan-European Agile model (tribes, squads, chapters)
- Monitored production health across stability (Crashlytics), feature performance (Google Analytics), and system traces (HoneyComb); crash-free session rate improved from 98.7% to 99.6%
- Started the app-wide migration to SwiftUI and contributed to the shared design system (including Objective-C components) consumed across 5 tribes
- Improved the OpenAPI definitions and code-generation pipeline, removing a long-standing source of client/server contract drift across SumUp's polyglot codebase
- Ran technical spikes (e.g. a rich-text editor via quill.js in a WebView as a lighter alternative to a native build) and provided feasibility feedback in iPhone and iPad design reviews

Dispo — Los Angeles, US
Founding Engineer — *Jan 2022-Feb 2023*

- One of the first technical hires; built and owned the real-time image-processing system and the camera-effects pipeline that became Dispo's product surface, including the release that grew daily-active count from 50,000 to 200,000
- Designed an asset-caching mechanism using async/await that reduced p99 image-load latency by 40% and cut memory usage by 30% at production scale
- Re-architected the application from The Composable Architecture to MVVM with SwiftUI, Combine, and async/await
- Ran the AB-experimentation infrastructure for new camera effects across 15 live experiments; prototyped disparity-based depth-measuring techniques for 3D effects, 3 of which shipped to production

Netguru — Poznań, PL
Junior iOS Developer → iOS Developer → Research and Development Lead — *Oct 2019-Jan 2022*

- Led the iOS Research and Development chapter: set the technical research agenda, evaluated emerging client problems against feasibility, and advised on architecture across 5 concurrent client projects
- Designed and shipped a cross-platform design system used by 8 apps across Android, iOS, and Flutter, with documentation, screenshot tests, and UI tests as first-class artefacts
- Architected a voice-assistant CRM application built on SwiftUI, Combine, on-device speech transcription, and a serverless backend over WebSockets — Netguru's first on-device ML production deployment
- Identified and patched 3 critical Firebase Authentication security vulnerabilities in a client IoT product
- Coordinated AR virtual try-on R&D and curated the iOS-trends review process for the chapter

OPEN SOURCE

wisent — github.com/wisent-ai/wisent

- Co-author and maintainer of the reference representation-engineering library for text LLMs: 342 GitHub stars, 32 forks, over 200,000 downloads, latest release v0.11.43 (2026-05-18), continuous releases since 2024

SwiftOLED & SwiftGFX — open source

- Two Swift libraries developed while tinkering with Raspberry Pi and SwiftOnArm: one for drawing pixel buffers on external OLED displays, the other a graphics library capable of rendering text to a pixel buffer. Both ship with READMEs and received external contributions

SKILLS, ACTIVITIES & INTERESTS

ML infrastructure: MLX (on-device LLMs, image, TTS), HuggingFace transformers, contrastive-pair pipelines, per-layer activation extraction, capability + safety eval harnesses
Cloud, serverless & backend: AWS (Lambda, EventBridge, EC2, S3, CloudFront), Firebase, Python, FastAPI, REST, WebSockets, gRPC, GraphQL, OpenAPI
CI/CD, observability & distribution: GitHub Actions (incl. self-hosted runners), Bitrise, Crashlytics, Google Analytics, Amplitude, Firebase Remote Config, HoneyComb, App Store Connect, TestFlight, Apple Business Manager, Jamf
iOS & tooling: SwiftUI, UIKit, async/await, Combine, RxSwift; Git/Git Flow, Swift Package Manager, CocoaPods, Carthage, Mint
Languages: English (fluent — IELTS C1)