

Norman Thomas.

Head of Engineering · Hands-on Full-Stack- & KI-Entwickler

job@norman.works

norman.works

github/norman-thomas

in/norman-thomas

München, Deutschland

Head of Engineering, der nie aufgehört hat zu bauen. Ich führe Teams und verantworte die technische Richtung — von der Cloud-Architektur bis zur letzten Zeile Frontend — und ich liefere: Datenprodukte, interne Tools und inzwischen KI-Agenten.

Offen für Neues · Festanstellung ab Oktober 2026 · Freiberuflich früher — zunächst Teilzeit, flexibel · Remote / Homeoffice

ERFAHRUNG

Head of Engineering Open Publishing GmbH · München, Deutschland

May 2019 – Present

Ich führe bis zu 12 Ingenieure (Entwicklung, Infrastruktur/Ops, QA) für die B2B-Software von Open Publishing für Verlage — und bleibe dabei hands-on. Ich setze die Architektur und verantworte den Wechsel in die „neue Welt“: Cloud, Microservices und ein modernes Frontend.

- Verantwortete den Übergang von einem teils-legacy C++-Monolithen zu AWS, Python-Microservices und TypeScript/React — eine offenere, wartbarere und erweiterbarere Plattform, die auch den Monolithen selbst schlanker und anpassbarer macht.
- Services auf dem neuen Stack gehen mit drastisch schnelleren Deploy- und Feedback-Zyklen live — enge need → build → verify → deploy-Schleifen und eine deutlich bessere Developer Experience als das Warten auf lange Monolith-Builds.
- Führe und mentore das Team und bringe Ingenieure — Junior wie Senior — in Cloud/AWS, Data Engineering, React und Python weiter.
- Solche kundenseitigen Funktionen — selten eine allein, sondern das Gesamtpaket — haben wiederholt geholfen, neue Verlagskunden zu überzeugen und zu gewinnen.

Catalog Analytics — Von Grund auf gebaut — hilft Verlagen, Katalog- und Metadatenprobleme zu erkennen, bevor Produkte in die Regale kommen. AWS-serverless (Lambda, API Gateway, DynamoDB, S3, CloudFormation) in Python/FastAPI, mit mehrsprachigem React/TypeScript-Frontend.

Financial Reporting Dashboard — Gibt Verlagen Klarheit über ihre Finanzen — flexible Abfragen, Filter, Visualisierungen und gespeicherte Favoriten, auf einem Redshift-Backend.

Mailing Tool — Versendet Leseexemplare an Journalisten und Rezensenten für Presse und Rezensionen.

Python · FastAPI · TypeScript · React · AWS · Microservices · DynamoDB · SNS · SQS · CloudFormation · GraphQL

Senior Full-Stack Developer Open Publishing GmbH · München, Deutschland

Jun 2013 – May 2019

Senior Full-Stack-Entwickler in einem verteilten, agilen Scrum-Team — volle Verantwortung von Konzept und API-Design über Implementierung und Tests bis zum Deployment. Pionier des ersten Microservices und der frühen ML-Arbeit des Unternehmens.

- Den ersten Microservice des Unternehmens (In-Book Marketing) end-to-end gebaut — die Pionierarbeit, die zur Vorlage für jeden weiteren Service wurde.
- GraphQL in die Projekte des Teams eingeführt — bei Entwicklern wie Kunden gut angekommen.
- NoSQL (MongoDB, AWS DynamoDB) dort eingeführt, wo relationale Modelle schlecht passten.
- Frühe Data-Science- und ML-Studien an Produkt- und Preisdaten durchgeführt.

In-Book Marketing — Der erste Microservice des Unternehmens — end-to-end gebaut (Python, Flask, GraphQL, MongoDB; React-Frontend; voll getestet; auf Docker Swarm deployed), mit einem AWS-nativen Prototyp (Lambda, DynamoDB, S3). Die Pionierarbeit, die jedem weiteren Service den Weg bereitet hat.

Book Pricing & Sales Analysis — Eine Data-Science-Studie dazu, welche Preise für Bücher/E-Books am besten funktionieren — geclustert nach Genre, Erscheinungsalter u. a. für den DACH-Markt; die Ergebnisse wurden Verlagen auf einer Buchmesse präsentiert. (pandas, scikit-learn, Keras, PyMC3)

Python · JavaScript · C++ · Flask · GraphQL · React · MongoDB · MySQL · Docker · pandas · scikit-learn · Keras

Full-Stack Developer GRIN Verlag GmbH · München, Deutschland

Dec 2009 – Jun 2013

Full-Stack-Entwickler in einer Phase rasanten Feature-Wachstums auf GRINs Publishing-Plattform — die Arbeit, die es ermöglichte, die Software als SaaS zu white-labeln, woraus Open Publishing entstand.

- Eine schnelle Welle von Full-Stack-Feature-Erweiterungen (C++ und JavaScript) auf der zentralen Publishing-Plattform geliefert.
- Diese Grundlage ermöglichte das White-Labeling der Software als SaaS-Produkt für andere Verlage — der Keim, aus dem Open Publishing wurde, später als eigenes Unternehmen gegründet und ausgegliedert.

C++ · JavaScript · MySQL

Das EJJE (英和和英辞典) Englisch-Japanisch-Wörterbuch end-to-end gebaut — Frontend und Backend.

- Das EJJE Englisch-Japanisch-Wörterbuch end-to-end gebaut, inklusive eines Parsers für ein proprietäres Binärformat im Backend.
- Bis heute weitgehend unverändert im Einsatz.

Java EE · Servlets · JSP · Tomcat · PostgreSQL

PRODUKTE

Ergo

Ein transparenter KI-Agent, dem man beim Denken zusehen kann.

Launch ~Aug. 2026

Ein KI-Agent in Kommandozeilen-Ästhetik — jedes Token, jeder Cent, jeder Tool-Call sichtbar. Modell-agnostisch: OpenRouter, OpenAI oder Anthropic. Die Tool-Calling-Ebene habe ich entworfen und implementiert.

AI Agent · LLM · Tool-calling · OpenRouter · TypeScript · ergo-ai.app

Papercut

Deine Briefpost, digitalisiert — die KI liest mit.

Live

Fotografiere einen Brief, den Rest macht die KI — eine Zusammenfassung, die wichtigen Aktionen (bezahlen, anrufen) und ein Live-Dashboard deiner Post. Rechnungen werden zu scannbaren Zahlungs-QR-Codes (kein Tippen von Empfänger, IBAN oder Verwendungszweck), Fälligkeiten werden nachgehalten, Werbung wird stummgeschaltet, aber nie verworfen, und alles — auch Steuerbriefe — ist per Multi-Facetten-Suche auf iPhone, iPad oder Mac griffbereit. Gegen Deutschlands Obsession mit Papier.

AI · OCR · iOS · macOS · Payments · Search · papercut.ink

Walko

Ein lokaler Guide in deiner Tasche — null Planung.

Landingpage live · App-Launch ~Aug. 2026

Entdeckt interessante Orte und bringt sie in eine maßgeschneiderte, praktische Route — und berücksichtigt dabei gleichzeitig verfügbare Zeit, Öffnungszeiten, Distanz und Nähe, Wetter, deine Interessen sowie wie beliebt oder einzigartig ein Ort ist.

AI · Travel · LLM · Mobile · OpenRouter · walko.guide

What Was I Thinking

Halt einen Gedanken fest, bevor er weg ist — direkt vom Handgelenk.

In Konzeption

Ein Tipp auf der Apple Watch, und du sprichst. Die App transkribiert das Gesagte, kategorisiert es und gruppiert es mit verwandten Gedanken — und erinnert dich abends daran, die vielversprechenden auszuarbeiten, wenn du magst. Kein App-Gesuche, keine Idee geht verloren, nur weil Tippen zu langsam war.

iOS · watchOS · AI · Voice · Transcription · whatwasithinking.app

KENNTNISSE

Programmiersprachen	Python · TypeScript · JavaScript · C++
Frontend	React · TypeScript · Astro · CSS
Backend & APIs	Microservices · GraphQL · REST · FastAPI · Flask
Cloud & Infra	AWS Lambda · S3 · DynamoDB · EC2 · API Gateway · SNS · SQS · CloudFormation · Serverless
Daten	Pandas · NumPy · scikit-learn · Data mining · MySQL · MongoDB · InfluxDB
KI / LLM	LLM integration · Agents & tool-calling · RAG · LangChain / LangGraph · OpenRouter / OpenAI / Anthropic
Praxis	Architecture · Team leadership · Mentoring · Testing (Unit, TDD, BDD) · Agile/Scrum

ZERTIFIKATE

Früh in ML/KI — die deeplearning.ai-Spezialisierung und angewandte Data Science bereits 2017–2018 — und weiterhin dran: aktuelle IBM-Kurse zu Agentic AI, RAG und LangChain/LangGraph (2026), die direkt in Ergo einfließen.	
Building AI Agents and Agentic Workflows Specialization	IBM · Jun 2026
Fundamentals of AI Agents Using RAG and LangChain	IBM · Apr 2026
Bayesian Statistics: From Concept to Data Analysis	Coursera · 2018
Applied Text Mining in Python	Coursera · 2018
Neural Networks and Deep Learning	deeplearning.ai · 2017
Improving Deep Neural Networks	deeplearning.ai · 2017
Structuring Machine Learning Projects	deeplearning.ai · 2017
Machine Learning (Andrew Ng)	Stanford · 2017

SPRACHEN

Deutsch (Muttersprache) · Französisch (Muttersprache) · Englisch (Fließend) · Kreolisch (Konversationssicher) · Japanisch (Konversationssicher) · Chinesisch (Grundkenntnisse) · Griechisch (Grundkenntnisse) · Persisch

(Grundkenntnisse)

AUSBILDUNG

Informatik (Diplom) Universität Kassel · Kassel, Deutschland

Sep 2001 – Mar 2007

Note 1,3 — entspricht US-GPA 4,0 — Ein mobiles Authentifizierungssystem mit ereignissynchronen Tokens.

Hübner-Stipendium Tongji-Universität · Shanghai, China

Sep 2005 – Apr 2006

Studium der chinesischen Sprache und Kultur.