Danni **Dromi**

Independent Data Scientist · Computer Science MSc.

Persillevej 44 - 2300 Copenhagen, Denmark

| | hello@finechariot.com | www.finechariot.com | danni-dromi

Data Scientist with 9 years industry experience working with AI, data and Software. Excels at handling data science projects from a full-stack approach. Experienced in building AI based solutions while communicating and understanding client needs and expectations. A creative problem solver with a passion for everything AI.

Past Project Experience

Semantic Search on Biology Research Documents

Client: Novo Nordisk

2023 - 2024

Role: Senior Data Scientist

- Integrate AI to an existing internal data driven platform
- Investigate usecase for using AI in searching research literature
- Understand stakeholder and user needs and how these can be managed by AI
- Analyse user tracking data to understand user behavior
- Determine best models and approaches. Discuss these with stakeholders.
- Extract and transform data for training a custom embedding LLM
- Fine tune model for client specific data
- Deploy several prototypes for internal and external user testing
- Communicate continuously with stakeholders and team leads
- Presented internally several times to raise AI literacy in broader team

Methods:

Semantic Search, Data Search, Large Language Models, NLP, Machine Learning, Data Modeling, Stakeholder Management, Client Presentation

Technologies:

Amazon Web Services, Sentence Transformers, Open Search, BERT, Azure Platform, Python, Polars, Streamlit, Neo4J

Term Data Analysis

Client: Teknologisk Institut

2024 - 2024

Role: Senior Data Scientist

- Perform analysis of client Term Data from the last 5 years
- Investigate performance of various courses.
- Discover trends and patterns
- Present results to client and deliver written report of findings

Methods:

Data analysis, Data modelling, Client Presentation

Technologies:

Python, Polars, Seaborn, Plotly

SkoleGPT

Client: Københavns Professionshøjskole

2023 - 2023

Role: Senior Data Scientist

- Prototype project to build a ChatGPT style chatbot for use in danish school system
- Focus on GDPR, Open Source, Compliance, without logging or tracking
- Worked with client to sketch and understand an MVP solution
- Researching Open Source Large Language Models to find best candidate for project
- Deployed hosted model with scaling
- Implemented front-end solution based on existing Open Source framework
- Further development is planned
- Project available at https://skolegpt.dk/

Methods:

Chatbots, Large Language Models, GDPR, Open Source, NLP, Machine Learning, Data Modeling, Prototyping, Full Stack

Technologies:

Azure Platform, Hugging Face, Llama 2, Python, Pandas, NextJS

Automated Fact Checking Study

Client: Amesto NextBridge

2022 - 2023

Role: Senior Data Scientist

- A preliminary study on how current AI models & tools can support fact checking organisations
- Collaborated with Amesto NextBridge on a delivery for their client Faktisk.no (Norwegian factchecking medium)
- Study and understand current processes in fact checking
- Analyze and understand cutting edge methods for using AI in fact checking from both the academic and professional world
- Map out all related technologies
- Develop and measure POC models & their performance
- Supported client (Amesto NextBridge) on building slidedecks & handover material
- Collaborated with client (Amesto NextBridge) on presenting the outcome of the project to their client (Faktisk.no)
- Currently working on publishing a research paper with client on project findings

Methods:

Fact Checking, Claim Detection, Claim Validation, NLP, Machine Learning, Exploratory Data Analysis, Academic Study, Client Management, Data Modeling

Technologies:

Azure Platform, Python, SQL, Pandas, OpenAI, BERT, PowerPoint

Food to Nutrition Matching

Client: Unikk.me

2022 - 2022

Role: Senior Data Scientist

- Develop a product for matching grocery store items to a nutritional database
- Analyze input data to understand and detect patterns and types of food items
- Map out and investigated multiple AI based models and hypotheses
- Continuously report on results and findings to clientside stakeholders
- Delivered a functional model that could easily be integrated in their existing Azure framework
- Demonstrate further methods of improvement to the client

Methods:

Data Analysis, Data Modeling, Data Visualization, NLP, Machine Learning, Stakeholder Management, Clustering

Technologies:

Azure Platform, Python, SQL, Pandas, Jupyter Notebooks, SpaCy, GIT

Modelling Data Architecture & Flow

Client: Unikk.me

2022 - 2022

Role: Senior Data Scientist

- Map out and specify the data model for client product under development
- Create a clear overview of how data flows through product stages
- Continuous dialog with client stakeholders to extract requirements for data
- Specify data columns and requirements at different product stages
- Create multiple data architecture diagrams and data structure tables based on client requirements
- Present findings to client over a series of workshop-style meetings

Methods:

Data Analysis, Data Modeling, Data Architecture, Stakeholder Management, Data Presentation

Technologies:

Python, SQL, Pandas, Jupyter Notebooks

AI Lead Scoring - Multiple deliveries

Client: GroupM

2020 - 2022

Role: Data Scientist

- Predict users likelihood to purchase from client on-site behavior
- Create custom ID matching and pull visitor behavior data from client website using Google Tag Manager
- Create ETL pipelines for continuously pulling data to GCP
- Perform data exploration and data modeling on gathered data to analyze and detect patterns and signals
- Build predictive models using BigQueryML
- Report to relevant stakeholders on progress & performance of solution
- Deliver and handover audiences on predefined advertisement platforms (Google DV360, GoogleAds, Facebook, etc.)
- Implemented on-site behavior based customer targeting strategies using Machine Learning.

Methods:

Machine Learning, Stakeholder Management, ETL Pipelines, Cloud Engineering, Exploratory Data Analysis, Data Modeling, Data Visualization

Technologies:

Google Cloud (BigQuery(ML), AutoML, Cloud Functions, Pub/Sub, Firestore, Data Studio, Stackdriver), Azure DevOps, Python, SQL, Pandas, NodeJS, GIT

ETL marketing attribution pipeline - Multiple Deliveries

Client: GroupM

2021 - 2022

Role: Data Engineer

- ETL integration for a marketing attribution dashboard - Multiple Deliveries
- Build a flexible ETL framework for pulling data from multiple sources
- Engineered for robustness and ease of use
- Fast deployment and easy to customize
- Running correspondence with project stakeholders regarding supporting additional platforms
- Data delivered in cleaned state in BigQuery in GCP

Methods:

Data Engineering, Cloud Engineering, Stakeholder Management, Cloud Architecture

Technologies:

Google Cloud (BigQuery, Cloud Functions, Pub/Sub, Firestore, Stackdriver), Python, SQL, Git

Sellout Data Pipeline

Client: GroupM

2021 - 2022

Role: Data Engineer

- Worked for a client of GroupM as part of a team to build a unified pipeline for ingesting sellout data
- Data arrived from multiple sources using different methods (email, FTP-server, self-service platforms, etc)
- Built ETL pipeline for ingesting massive batch datasets
- Contributed on ETL Framework for flexibly ingesting data from various email sources
- Helped client migrate an existing solution to a cloud hosted environment
- Partook in workshops and presentations to handover solutions and knowledge obtained during the implementation phase to client

Methods:

Data Engineering, Client Management, Client Mediation, ETL Pipelines, Cloud Architecture

Technologies:

Google Cloud (Pyspark, BigQuery, Cloud Functions, Compute (VM's), Stackdriver), Azure DevOps, Python, Pandas, SQL, Git

Activation Framework

Client: GroupM

2020 - 2021

Role: Data Engineer

- Worked for a client of GroupM as part of a team to build pipelines for uploading datasets to activation platforms
- Partook in client process meetings (Daily standup, sprint planning, scoping, etc.)
- Collaborated with client-side engineers
- Extended upon existing client framework and client methodologies
- Performance tested and reported on built solutions

Methods:

SCRUM, Data Engineering, ETL Pipeline, Client Management

Technologies:

Microsoft Azure (Databricks, DataLake Storage, Cosmos DB), Pyspark, Apache Spark, Python, SQL, Git

Video Analysis Framework

Client: GroupM

2020 - 2020

Role: Data Scientist

- Worked in collaboration with two MIT students to build a pipeline for analyzing a commercial video
- Split a commercial into separate cuts by Machine Learning model
- Analyze each cut in various steps based on still image, audio & detected text
- Each analysis done using different pretrained Machine Learning model
- Partial results & final results written to GCP
- Project used experimentally for various use cases (eg. analyzing an episode of the DR2 show “Deadline”)

Methods:

Machine Learning Engineering, Cloud Engineering, Data Science

Technologies:

Google Cloud (Storage (buckets), BigQuery, AutoML, Stackdriver, Compute (VMs), Python, Shell Script, Keras, Git

Docunet

Client: Paperflow

2019 - 2020

Role: Machine Learning Engineer

- Worked on a Machine Learning model for digitally flattening creased documents
- Method based on published academic research
- Built code for creating and synthesizing training set
- Developed model architecture based on described research

Methods:

Machine Learning Engineering, Data Science

Technologies:

Tensorflow, Python, Numpy, Pandas, SQL, Scipy, Convolutional Neural Networks (CNN), Deep Neural Networks (DNN)

Catalog Classifier

Client: Paperflow

2018 - 2019

Role: Machine Learning Engineer

- Classify and match invoices by creditor to a customer provided catalog
- Build infrastructure for gathering data from multiple data sources
- Clean-up and structure data
- Experiment with model architectures
- Implement system of Machine Learning models
- Dynamically predict and retrain models using GCP
- Report and correspond with customers and stakeholders on performance of solution and ad-hoc problems

Methods:

Machine Learning Engineering, Data Science, Stakeholder Management, Cloud Engineering

Technologies:

Google Cloud (AI Engine, Storage), Python, Keras, Tensorflow, Numpy, Pandas, SQL, Scipy, Convolutional Neural Networks (CNN), Deep Neural Networks (DNN), Transfer Learning

Development and maintenance

Client: Paperflow

2018 - 2020

Role: Back-End Engineer

- Assisted in various development tasks
- Improved functionality of existing back-end solution
- Handled Ad-Hoc bugs and issues
- Deployment and scaling to production environments

Methods:

Software Engineering, SCRUM

Technologies:

Python, Kubernetes, Docker, PostGreSQL, Git

Education

UCPH (University of Copenhagen)

MSc. in Computer Science — *Nov. 2014*

- Elective Project: Hybridization of mesh-free and mesh based simulation methods
  developed using MATLAB, grade: 10
- Studied a semester abroad (aug. - dec. 2013) at the *University of British Columbia*, Vancouver - Canada
- Master's Thesis: The Use of Moving Meshes for Segmentation and Registration of 2D Medical Imaging
  developed using C++, grade: 12

UBC (University of British Columbia)

Exchange Student — *Aug. 2013 - Dec. 2013*

- I studied a semester abroad, at UBC in Vancouver - Canada

UCPH (University of Copenhagen)

BSc. in Computer Science — *Jun. 2012*

- BSc. Project: Using Compounds of Convex Shapes in Interactive Rigid Body Simulation
  developed using C++, grade: 12

Frederiksberg Gymnasium

Mathematical Student — *Jul. 2007*

Talks

COMPUTERWORLD - AI Business Excellence Day — *Virtual*

Safe Applications of AI on your own Data — *Dec. 2023*

I gave a talk on how businesses can apply AI LLMS on their own data without compromising security. I also showcases uses for this technology.

Driving IT — *IDA - Copenhagen*

Wilder Generative AI — *Nov. 2023*

I gave a talk on generative AI and what progress we will see in the near and distant future.

ComputerWorld Cloud Festival — *Copenhagen*

3 Impressive Cases of Applied AI — *Sep. 2023*

I gave a talk together with Søren Christian Søndergaard Poulsen showcasing some of the most impressive usecases of AI both in Denmark and internationally.

Internal Presentation — *UFST (Skat) - Copenhagen*

SkoleGPT Presentation — *Sep. 2023*

I gave a talk at UFST on my work in developing SkoleGPT as well as the experiences I've gained in working with cutting edge Large Language Models.

Driving IT — *IDA - Copenhagen*

Deep Generative Models — *Nov. 2022*

I gave a talk on Generative AI, its history and progress (right before ChatGPT came out).

PyData Copenhagen — *Alm Brand - Copenhagen*

8 years of Deep Generative Machine Learning — *Sep. 2022*

I gave a talk on the exciting and rapidly progressing history of Deep Generative ML and also showcased some of my own creative work.

Additional

Languages Danish (mother tongue), English (high professional level), German (still learning)

Communication Excellent written and verbal communication as well as presentation skills
Experienced working in a client facing role

Hobbies

Yoga	Been actively practicing for the last 3 years
Motorcycling	Proud owner of a 2010 Honda Hornet
Diving	PADI Advanced Open Water certified. Completed over 30 dives.

Skill Matrix

Skill Name/Category	Level	Experience	Last Used
Industry			
Pharmaceuticals	★★★☆☆	1 year1	2024
Education	★★★☆☆	1 year	2024
Marketing	★★★★☆	2 years	2022
FinTech	★★★★☆	4 years	2020
Data Science & Advanced Analytics			
Data Modelling	★★★★★	5 years	2024
Large Language Models	★★★★★	5 years	2024
Natural Lanauge Processing	★★★★★	5 years	2024
Semantic Search	★★★★☆	1 years	2024
Conversational Language Models (Chatbots)	★★★★☆	1 years	2024
Hugging Face Platform	★★★☆☆	1 years	2024
Data Analysis	★★★★☆	4 years	2024
Sentence Transformers (SBert)	★★★★☆	1 years	2024
OpenAI Services	★★★★☆	1 years	2024
Machine Learning	★★★★★	5 years	2024
Pandas + Polars	★★★★★	4 years	2024
Data Visualization	★★★★☆	3 years	2023
Keras	★★★☆☆	2 years	2020
Development Language & Technology			
Python	★★★★★	7 years	2024
Node.js/Javascript	★★★★★	4 years	2023
Git	★★★★★	10 years	2024
Docker	★★★☆☆	3 years	2024
Database			
SQL	★★★★★	10 years	2024
Neo4J	★★★☆☆	1 years	2024
PostgreSQL	★★★★☆	5 years	2020
Redis	★★☆☆☆	2 years	2020
Platform			
Google Cloud Platform (GCP)	★★★★★	4 years	2023
Amazon Web services (AWS)	★★★★☆	3 years	2024
Microsoft Azure	★★★★☆	2 year	2024
Linux	★★★★★	12 years	2024
Development Process			
Continuous Integration/Continuous Deployment (CI/CD)	★★★★☆	5 years	2022
Testing (Unit + Integration)	★★★★★	6 years	2022
Internal QA (Code Review)	★★★★☆	4 years	2021
Estimation	★★★★☆	4 years	2024
Communication			
Presenting	★★★★★	2 years	2024
Stakeholder Management	★★★★☆	2 years	2024
Client Communication	★★★★☆	3 years	2024
Workshop Facilitation	★★★★☆	2 years	2024