Rapport de stage de fin d'études

À retourner au plus tard ***1 semaine avant la date de la soutenance***

	Rapport de Stage **Ing5 - 3ème année Cycle Ingénieur** **Promotion 2025**
Nom : Meyer **Prénom :** Jacques	**Majeure :** ☒ DATA & IA ❒ CREATE ❒ DEF ❒ DT& I ❒ EE ❒ FIN ❒ SANTE ❒ SI/CYB/CLOUD ❒ SE/VCA

Entreprise d'accueil

Nom : Cognizant

Adresse : 280 Bishopsgate London EC2M 4AG

Adresse du lieu de stage si différent :

Engagement de Confidentialité : ❒ oui ☒ non

Reçu le .02/08/2025

Rapport Confidentiel à remettre au tuteur de stage à l'issue de la soutenance : ❒ oui ☒ non

____

Pénalités Observées :

Description de la mission

The objective was to design and develop a prototype application that generates personalized weather forecast videos based on user-selected dates and locations. These videos include visual elements such as maps and graphs, natural voice narration, and a virtual avatar, and cover a 72-hour forecast window. The entire solution was built using AWS services, ensuring scalability and performance.

Cognizant UK
280 Bishopsgate London
11/07/2025

Internship Report

Personalized Weather Forecast Video Generator using GenCast and AWS

April to September 2025

Table of Contents :

Acknowledgements

First and foremost, I would like to express my sincere gratitude to the Artificial Intelligence and Analytics (AIA) team at Cognizant UK for welcoming me and supporting me throughout this internship experience. Working within such a dynamic and innovative environment has been incredibly enriching, both professionally and personally.

I would especially like to thank Neel Savani, my internship supervisor, for his guidance, encouragement, and trust throughout the project. His insights and availability were instrumental in helping me navigate the technical and strategic aspects of the work. His mentorship played a key role in the successful development of the project and in my growth as a future engineer.

I also wish to thank the broader Cognizant team for their collaboration and openness, which allowed me to integrate quickly and contribute meaningfully to the project. Their feedback and support helped me refine my ideas and deliver a solution aligned with real-world client needs.

I am also grateful to my school for enabling this internship opportunity and for the continuous academic support provided during this period.

Finally, I would like to thank my family and friends for their encouragement and support throughout this experience, which helped me stay focused and motivated.

Summary

As part of my fifth year at ECE Paris, I completed a six-month internship at Cognizant UK, working as a Programmer Analyst. My project focused on exploring and showcasing the capabilities of GenCast, a machine learning-based weather prediction model developed by DeepMind.

The objective was to design and develop a prototype application that generates personalized weather forecast videos based on user-selected dates and locations. The application should include video visual elements such as maps and graphs, natural voice narration, a virtual avatar, and cover a 72-hour forecast window. The entire solution was to be built using AWS services, ensuring scalability and performance.

In addition to the technical development, the project involved evaluating the performance of different GenCast model configurations, identifying the best real-time data sources, and exploring ways to increase the model’s temporal resolution from 12-hour to hourly forecasts. The final deliverable was intended to serve as a client-facing demo, showcasing how GenCast can be adapted to meet specific industry needs, such as renewable energy forecasting for offshore wind farms.

1. Introduction

My name is Jacques Meyer. I am 22 years old and currently in my fifth and last year as an engineering student at ECE Paris. I chose this generalist engineering school because of its broad range of majors, which allowed me to explore various fields before choosing to specialize. Over the course of my studies, I developed a strong interest in computer science, particularly in the areas of artificial intelligence and data science, in which I had the chance to specialize during my final year.

I consider myself a rigorous and adaptable person, capable of working both independently and as part of a team. I am curious by nature and passionate about discovering new technologies and cultures, which is why I also enjoy traveling and learning from diverse environments. My professional goal is to become an engineer in the digital field, working on innovative and impactful projects.

To gain experience in AI and data, I actively searched for an internship aligned with these interests. After several applications and opportunities, I interviewed with Cognizant Technology Solutions. This gave me the chance to move to London and to join Cognizant UK for a six-month internship.

As part of this internship, I worked as a Programmer Analyst on a project centered around GenCast, a machine learning-based weather prediction model developed by DeepMind. The objective was to design a prototype that generates personalized weather forecast videos based on user-selected locations and dates. The project solution should integrate AWS services for hosting and scalability, and include visual elements, voice narration, and a virtual avatar to deliver a 72-hour forecast.

This report outlines in more detail the work I carried out during my internship at Cognizant. Following an introduction to the company, the report gives an overview of the project and its context, then a detailed description of my missions and the technologies used. The report concludes with a reflection on the experience and the skills I developed throughout the internship.

2. The Company

2.1 Cognizant as a Company

Cognizant Technology Solutions is an international IT services, consulting, and business process solutions company headquartered in New Jersey, USA. It's widely recognized for its work in digital transformation, IT consulting, and outsourcing. In the 2025 Fortune 500 list, Cognizant ranked #217 overall and was the third-highest among pure IT consulting firms. As of 2024, Cognizant generates more than $19 billion in annual revenue and employs over 336,000 people worldwide, with a strong presence across North America, Europe, and Asia.

The company started out in 1994 in Chennai, India, as an internal tech unit for Dun & Bradstreet, and became an independent business in 1996. Today, the company's services are organized into three main areas: Digital Business, Digital Operations, and Digital Systems & Technology. These cover everything from customer experience design and digital product development to AI-powered automation and IT infrastructure. Cognizant works with clients across a wide range of industries, including finance, healthcare, retail, manufacturing, and the public sector.

Its portfolio includes cloud services, AI and analytics, enterprise applications, digital engineering, software development, and cybersecurity. All of these are aimed at helping clients adapt to digital change, stay competitive, and improve how they operate.

2.2 My Department

During my internship, I was part of the Artificial Intelligence and Analytics (AIA) department, which sits within Cognizant Digital Business (CDB). The team focuses on using data science, machine learning, and AI to help clients make smarter, data-driven decisions.

In the UK, Cognizant AIA works with a wide range of clients across both private industry sectors such as finance, healthcare, retail, as well as in the public sector. Throughout my time there, I got to observe projects involving well-known clients such as Novartis, Aston Martin Formula One, the Football Association (FA), and Sanofi. I also saw how Cognizant supports digital transformation efforts in the public sector, including work with the UK Department of Justice to modernize systems and improve services.

3. Corporate Social Responsibility (CSR)

3.1 Introduction

For just over three decades, Cognizant has been helping organizations around the world navigate waves of technological change. Entering the challenging era of Artificial Intelligence (AI), the company continues to lead —leveraging innovation to drive business transformation while creating long-term value for clients, communities, and the planet.

As Cognizant's Chief Executive Officer stated in the 2024 Sustainability and Corporate Citizenship Report:

"We see AI as a generational opportunity, and are embracing its potential with confidence. At the same time, we remain committed to developing and deploying AI responsibly—advancing both business and society while managing risk and environment."

Cognizant's strategy is based on a commitment to people, innovation, and sustainability. These values are embedded across the company's operations and corporate social responsibility (CSR) efforts and can be divided into two broad areas: People and Communities, and Environmental Sustainability.

3.2 People and Communities

By investing in its people and the wider community, Cognizant shows a commitment to social impact, innovation, and long-term, sustainable growth. Many of the initiatives highlighted below strongly reflect the company's roots in India, with a significant focus on community impact and development in the region.

3.2.1 Skilling for the Future :

According to Cognizant's 2024 "New Work, New World" study, performed in collaboration with Oxford Economics, around 90% of jobs are expected to be affected by generative AI over the next decade. As part of its CSR efforts, Cognizant launched Synapse, a global initiative designed to equip both Cognizant employees and people and communities outside the company, with the tools and knowledge they need to succeed in this fast-changing environment. The goal is to positively impact one million lives by the end of 2026.

The program is built around five main pillars:

- Skills Accelerator: Focused on building a tech-ready workforce. This program has trained over 3,700 people globally and aims to upskill 200,000 job seekers through sponsored programs and partnerships.
- Apprenticeships & Internships: In collaboration with universities and workforce groups, Cognizant offers paid, hands-on learning opportunities. So far, 530 individuals—mainly in the U.S.—have taken part.
- Employee Skilling: Training staff in areas such as generative AI, cloud, and other emerging technologies. Training is delivered through partnerships with major tech companies including Google, Microsoft, AWS, Salesforce, SAP, and Oracle. Over 2,000 leaders have completed specialized gen AI learning paths and more than 173,000 company associates have been trained in these new skill areas.
- Technology Partnerships: Collaborations with companies like Microsoft and Google have expanded access to training. For example, 70,000 developers have been trained on GitHub Copilot, and 3,200 associates have completed Google AI training.
- Community Education: Through grants, mentoring, and volunteering, the company has supported over 164,000 individuals globally. More than 2,000 employees have contributed 20,000 hours to volunteer-led skilling projects. The goal is to reach 230,000 people by 2026.

In its first year, the Synapse initiative reached over 40% of its goal by delivering more than 400,000 impactful learning experiences. 164,000 Cognizant employees received training in at least one digital skill in 2024 alone.

As a paid intern, I was one of the fortunate beneficiaries of this program and several of its pillars.

3.2.2 Fostering Innovation:

Cognizant also fosters a culture of innovation. Employees are encouraged to submit ideas that can be developed either for internal use or in collaboration with clients, through the Bluebolt Program, an internal innovation engine powered by generative AI. This grassroots initiative supports fast experimentation and prototyping, offering resources like tutorials and dedicated creative spaces known as Bluebolt Garages.

Around 240,000 ideas have been submitted so far—over 47,000 of which have been implemented. These ideas are reported to have led to measurable improvements in service delivery, client satisfaction, and business performance.

In addition, Cognizant organizes regular hackathons and ideathons, to help spark creativity and collaboration across teams.

3.2.3 Learning:

Learning and development are a core business strategy for Cognizant. Operationally, from onboarding new hires to helping experienced professionals grow, there's an emphasis on continuous learning. This includes a range of digital training and leadership development programs to support both career progression and adaptability in a fast-changing tech landscape.

3.2.4 People Development and Management:

Cognizant aims to attract, develop, and keep top talent at every level of the company. There is a focus on talent development and creating a high-performance workforce.

Across all staff levels, there are programs in place to identify future leaders, with senior management involved and regular updates shared with the Board of Directors.

Some of the key initiatives include:

Customized career development: Career paths tailored to different stages — from starting out, to an internal move, or aiming for leadership roles. Promotion cycles happen twice a year.

Performance management: All full-time employees take part in a structured performance cycle that directly impacts bonuses, promotions, and salary reviews.

Role rotation and leadership job boards: To help employees explore new areas, build broader skills, and grow their network within the company.

Recognition programs: Celebrating achievements through both monetary rewards and peer-driven recognition, to build a culture of appreciation. In 2024, Cognizant introduced the Impact Awards to highlight outstanding contributions across teams and regions. The Spirit Award—the company's highest honor—recognizes exceptional performance and leadership.

3.2.5 Supporting Associate Wellbeing

Cognizant's Be Well program is a global initiative to support the overall wellbeing of employees and their families.It's designed to be flexible, helping people manage both work and life challenges more effectively.

The program is built around four pillars :

- Physical Wellbeing: Includes preventive care, telehealth, maternity support, cancer awareness, and fitness classes.
- Mental Wellbeing: Offers resources for managing stress, burnout, and mental health, including mindfulness tools, coaching, and awareness campaigns.

- Financial Wellbeing: Provides education on budgeting, investing, and retirement planning, plus access to financial advisors and discount programs.
- Life and Work Wellbeing: Supports work-life balance through flexible work policies, time-off programs, parenting resources, and employee affinity groups.

Some highlights from 2024 include:

- ❖ Over 400 wellbeing events were held, with more than 90,000 participants.
- ❖ Around 41,000 associates took part in physical and mental health challenges.
- ❖ The Mental Health Ally program now includes over 200 trained allies across 29 countries, offering confidential support in 26 languages.

3.2.6. Engagement and Retention

To keep engagement high and employees supported, Cognizant uses several strategies:

- Annual engagement surveys followed by action plans at both the company and team levels.
- Small group listening sessions with senior leaders to hear directly from employees.
- Ongoing engagement through training, mentoring, volunteering, and partnerships with external organizations.

The company also closely tracks voluntary attrition—especially in its Tech Services segment—to stay ahead of potential retention challenges and adjust its talent strategy as needed.

In 2024, the company employed over 57,400 new hires. Its annual People Engagement Survey, which gathered feedback from more than 231,000 employees, showed that Cognizant's engagement scores were above both global and IT industry benchmarks in most areas.

3.2.7. Social Impact

Cognizant's social impact strategy centres on technology as a tool for positive change. Through a mix of philanthropy, volunteering, mentoring, and education partnerships, the company works to support communities and help people prepare for the jobs of the future. These efforts are closely tied to the broader Synapse initiative, which aims to equip over one million individuals with digital skills by 2026.

3.2.8. Community Engagement and Volunteering

Cognizant encourages its employees to give back by sharing their time and skills with causes that matter to them. In 2024, more than 47,000 volunteers across the company contributed over 220,000 hours to a wide range of community-focused initiatives around the world.

- Global Philanthropy and Strategic Giving

In 2024, Cognizant awarded $15 million in grants to more than 130 organizations around the world. These investments are aimed at expanding access to tech education and helping more people build careers in the digital economy. Support of global nonprofits is managed through donor-advised funds and direct grants, with a focus on boosting economic mobility and strengthening community resilience.

- Cognizant Foundation India

The Cognizant Foundation in India, leads Cognizant's CSR efforts locally, focusing on inclusion, healthcare, and education at a national level. In 2024, The Foundation supported over 90 projects in partnership with 45 nonprofit organizations with initiatives such as Health4All, Sight4All, Care4All, and Support4All: (programs aimed to improve access to quality healthcare) ;STEAM4All and Tech4All: (Focused on promotion of education and digital skills in government schools, including young people with disabilities); and Excellence4All: (A program that supports access to higher education for underserved communities)

3.3 Environmental Sustainability

3.3.1 Introduction

Cognizant's environmental sustainability strategy is focused on cutting greenhouse gas (GHG) emissions, managing climate-related risks, and using resources responsibly across its global operations. To strengthen its capabilities in such areas as climate risk analysis, sustainability data management, and decarbonization planning, Cognizant also partners with companies like Microsoft, IBM, Salesforce, Tidal, CoolPlanet, and RS Metrics. At the same time, Cognizant is identifying opportunities to support its clients and communities in making responsible environmental transitions themselves .

3.3.2 Emissions Reduction

The company has set a path toward achieving net zero greenhouse gas emissions, backed by science-based targets that have been validated by the Science Based Targets initiative (SBTi).

Its goals are a 50% reduction by 2030 (compared to 2019 levels), and a 90% reduction by 2040, with any remaining emissions offset starting in 2030. By 2024, it had reduced total emissions by 52% compared to 2019. It also reported a 59% drop in emissions intensity, which now stands at 24.89 metric tons of CO_2 equivalent per million dollars of revenue.

To reach its net zero goals, Cognizant is focusing on six main strategies:

- Sourcing renewable electricity: In 2024, 46% of the company's global electricity use came from renewable sources, with a goal of reaching 100% by 2026.
- Reducing travel and commuting emissions: Business travel emissions were cut by 9% in 2024. Cognizant also operates a fleet of over 600 electric vehicles to support employee commuting.
- Engaging suppliers: 53% of suppliers in high-emission categories have committed to science-based emissions reduction targets.
- Improving energy efficiency: The company is using AI-powered HVAC systems and investing in green buildings—over 60% of its owned office space in India now meets LEED or similar standards.
- Sourcing high-quality carbon offsets: Cognizant is developing principles to guide its carbon offset purchases, ensuring they meet credible standards.
- Building a climate-aware workforce: In 2024, more than 15,800 employees across 35 countries completed climate training.

3.3.3 Energy Efficiency in Operations

Cognizant is working to improve energy efficiency across its facilities and IT infrastructure. One example is the use of AI-based chiller automation, which was first piloted in Chennai and is now being rolled out more widely. In terms of IT infrastructure, by 2024, 79% of its servers had been virtualized. Cognizant is also upgrading its data centers and moving more workloads to the cloud, achieving a Power Use Effectiveness (PUE) of 1.84 in its Chennai and Pune facilities.

3.3.4. Supply Chain and Travel Emissions

In 2024, about half of Cognizant's total emissions came from procurement. The company is now working closely with suppliers to improve emissions reporting and encourage science-based reduction targets. Travel made up 18% of emissions, mostly from air travel, and new data systems are helping identify and reduce inefficient travel practices.

Cognizant is also tracking emissions from remote work, which totaled 154,108 metric tons of CO_2 equivalent in 2024—a 9% drop compared to 2023.

3.3.5. Managing Climate Risks

Cognizant takes a proactive, science-based approach to managing climate-related risks and reducing its environmental impact. The company brings together governance, risk management, and day-to-day operations to address both the physical risks of climate change and the challenges of transitioning to a low-carbon economy. The company offers climate training to employees to build a climate-competent workforce. By the end of 2024, over 15,800 associates across 35 countries completed climate training.

3.3.6. Environmental Management and Nature Impacts

Cognizant runs an ISO 14001-certified Environmental Management System (EMS), which covered over 92% of its global facilities in 2024. The company's main nature-related impacts are linked to water usage and waste generation, particularly in its operations in India. Some examples of progress in reducing the company's impact in these areas are shown in the slides below:

Water Management:

Water data[1]	2023	2024
Total water withdrawal across all Cognizant sites[2] (in million liters)	741	890
Total water withdrawal (in liters per square meter)	344	430
Percentage of water withdrawal in high or extremely high baseline water stressed regions	62%	70%

We have undertaken initiatives at our owned facilities to help reduce water withdrawal, including:

- Harvesting of rainwater, which met approximately 5.6% of our water withdrawal requirements
- Recovery of condensate from Air Handling Units (AHUs), which contributed 1.2% to our water withdrawal need
- Reuse of all treated water for horticulture, toilets and cooling towers across some of our owned sites

Waste Management: Breakdown of waste in India (2024):

Waste data	2023 (tonnes)	2024 (tonnes)
Battery waste	348.44	218.55
Biomedical waste	0.28	0.23
Hazardous waste[4]	29.03	30.11
Non-hazardous waste[5]	587.62	784.15
E-waste (Operational)[6]	73.58	19.64
Total waste[7]	1,038.95	1,052.68
Waste reused or recycled[8]	97%	95%

In 2024, a total of 1,052.68 tonnes of waste was generated, with 96% of e-waste successfully diverted from landfill, over 80,000 laptops donated to communities in India, and a clear goal set to achieve zero e-waste to landfill by 2030.

3.3.7. Client Business Sustainability Solutions

Cognizant has a “Solving for Sustainability” group which offers strategy and advice as well as hands-on implementation of a range of tech-driven solutions to help businesses reduce their environmental impact, stay on top of regulations, and turn their climate goals into real, measurable actions. This includes support in nine key areas:

- Net Zero Energy Management
  Using IoT, AI, and automation to track and manage energy use in real time, helping clients hit their net zero targets.
- Sustainable Manufacturing and Operations
  Improving transparency and efficiency by tracking environmental and energy performance.
- Sustainability Reporting and Data Management
  Making it easier to collect and report data accurately, ensuring compliance and reliable carbon accounting.
- Sustainable Finance
  Helping clients manage climate-related financial risks and align their investments with net zero goals.
- Sustainable Supply Chain
  Encouraging transparency, ethical practices, and human rights due diligence across supply chains.
- Sustainable IT
  Cutting down IT-related emissions through green cloud solutions, energy-efficient data centers, and eco-friendly practices.
- Sustainable Products and Circular Economy
  Supporting circular economy efforts and product life cycle assessments using automation and data insights.
- Sustainability Strategy and Consulting
  Helping clients set long-term sustainability goals, identify key issues, and build actionable plans.
- Climate Resilience and Adaptation
  Using geospatial analytics and AI to assess climate risks and support decarbonization strategies.

3.3.8. Climate Training and Awareness

Finally, Cognizant is building a climate-competent workforce to support its sustainability goals. By the end of 2024, over 15,800 associates across 35 countries completed climate training.

My project aligns perfectly with Cognizant’s environmental sustainability goals, through its application to improving the effective use of renewable energy resources; predicting the emission and dispersion of toxic chemicals to protect environment and communities; and being a more energy-efficient solution for generating forecasts compared to currently available options.

4. My Team

4.1 Team Presentation

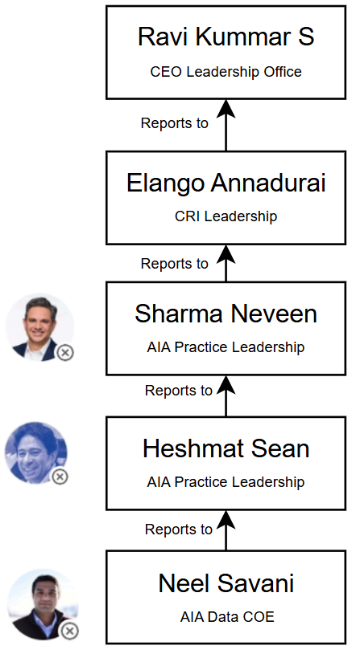
During my internship, I had the chance to work alongside highly skilled and experienced individuals, which made the experience especially enriching. In addition, each of them brought strong scientific and technical backgrounds, from which I could learn a lot, working with them day to day. In the next section, I'll briefly introduce the two people I collaborated with most closely :

Neel Savani, Director of the Gen AI Innovation Studio at Cognizant, has a multidisciplinary background that combines advanced scientific research with business and technology leadership. He holds a first-class MSci in Physics and a PhD in Space Physics from Imperial College London, where his doctoral work focused on the structure and propagation of coronal mass ejections—solar events that influence geomagnetic activity on Earth. He later completed an MBA at Imperial Business School, concentrating on strategy, innovation, marketing, and analytics.

Before joining Cognizant, Neel spent nearly seven years at NASA's Goddard Space Flight Center in the United States. There, he held roles as a Research Scientist and Media Product Manager, contributing to major space missions and leading teams in areas such as spacecraft engineering, data visualization, and virtual reality applications. His work included securing funding for next-generation missions and participating in international steering committees involving agencies like the UK Cabinet Office and the British Geological Survey.

Neel returned to London in 2020 and joined Cognizant as a Product Data Scientist for Apple. Today and since June 2024, Neel holds the position of AIA Data COE, where he leads the Gen AI Innovation Studio and manages a team focused on accelerating digital transformation for clients across sectors. His responsibilities include overseeing AI-driven product development, ensuring responsible AI practices, and aligning solutions with ESG standards. His career reflects a strong foundation in both scientific inquiry and practical innovation.

Neels Cognizant organisation diagram



Riccardo Bassiri, with whom I worked closely during my internship, has followed a career path that mirrors Neel Savani's in many ways—transitioning from a distinguished academic and research background to a leadership role in industry. Riccardo holds a PhD in Physics from the University of Glasgow and completed postdoctoral research at Stanford University, where he spent over a decade contributing to major scientific projects. Notably, he co-authored key publications related to the discovery of gravitational waves, a breakthrough that earned the LIGO team the Nobel Prize in Physics in 2017.

At Stanford, Riccardo served as Senior Research Scientist and Deputy Director of the Center for Coatings Research, leading multi-institutional collaborations focused on advanced materials for optical applications. His work combined deep technical expertise with strategic project management across global research teams.

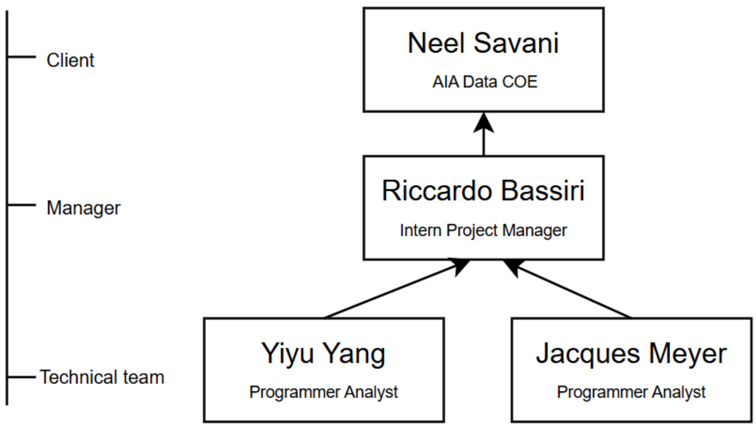
In 2024, Riccardo left academia to return to Europe and pursue new challenges in business and innovation. He is currently completing an MBA at Imperial College London, with a focus on entrepreneurship and venture capital. At Cognizant, he holds the role of intern project manager, contributing to AI and deep tech initiatives while building on his scientific and commercial experience.

4.2 Team organization

During my internship, the project team was structured to simulate a real-world client-consulting engagement. Neel Savani acted as the client, providing strategic direction and feedback. Riccardo Bassiri, serving as project manager, coordinated the team's efforts and ensured alignment with project goals. The technical team initially consisted of Yiyu Yang, a data science student at Imperial College London, and myself. Together, we were responsible for the design and development of the technical solution.

The objective was to deliver a functional prototype to a hypothetical client, following an Agile, iterative approach. We held weekly review meetings with Neel to present progress and gather feedback, complemented by more frequent internal working sessions. Midway through the internship, Yiyu had to step away from the project, leaving me solely responsible for the technical implementation. This shift required me to take full ownership of the development process and adapt quickly to ensure continuity and delivery.

My Team, organisation diagram



5. Project Context

5.1 Introduction: The Evolution of Weather Forecasting

For decades, weather forecasting has been one of the most computationally demanding scientific challenges. Traditional forecasts are generated using numerical weather prediction (NWP) models, which simulate the Earth's atmosphere based on physical laws—such as thermodynamics, fluid dynamics, and radiation transfer. These models require immense computing power, often running on some of the largest supercomputers in the world.

A prime example is the UK Met Office, which operates a supercomputer system valued at £1.2 billion, capable of performing 60 quadrillion calculations per second. These systems ingest hundreds of billions of weather observations daily and run models with millions of lines of code. The Earth is divided into a 3D grid, and the model calculates how weather variables evolve in each grid cell over time. However, the resolution of these grids—typically between 10 km and 28 km—limits the model's ability to capture small-scale phenomena like localized showers or mountainous terrain. Higher-resolution models, such as the UKV model with a 1.5 km resolution, offer better detail but are computationally expensive and limited to more developed regions like the UK, US and Europe.

Despite their accuracy and scientific rigor, traditional models are slow, often taking hours to produce forecasts, and require specialized infrastructure to operate.

5.2 The Rise of AI-Based Weather Models

In recent years, AI and machine learning have introduced a disruptive shift in the field of meteorology. Instead of simulating the atmosphere through physical equations, AI-based models learn patterns from vast amounts of historical weather data. These models are trained on decades of global weather records, enabling them to make predictions based on statistical correlations rather than physical simulations. Some of the most notable AI weather models include:GraphCast by Google DeepMind, AIFS by ECMWF (European Centre for Medium-Range Weather Forecasts), Aurora by Microsoft and Pangu-Weather by Huawei.

A big advantage of AI models is their speed. They can generate forecasts in under a minute on standard hardware, making them orders of magnitude faster than traditional NWP systems. In terms of accuracy, in recent evaluations, models like GraphCast, AIFS, and Aurora have even outperformed traditional models in certain metrics, such as predicting atmospheric pressure patterns.

The emergence of AI in weather forecasting raises critical questions for the future of the field: Can AI models replace traditional physics-based models, or will they serve as complementary tools? What are the risks and limitations of relying on data-driven models? How can we ensure trust, transparency, and robustness in AI-generated forecasts? And most importantly: Are these models ready for real-world deployment, or is further research and validation needed?

These questions are at the heart of ongoing research and experimentation — including the work I conducted during my internship at Cognizant with the GenCast model, which I will cover in the next section.

Source : https://www.bbc.co.uk/weather/articles/cwy6ykp7049o

5.3 My Role at Cognizant: Exploring the GenCast Model

During my internship at Cognizant, I worked as a Programmer Analyst on a project centered around GenCast, an AI-based weather forecasting model. My mission was to explore the capabilities and limitations of this emerging technology and assess its potential for future client applications.

The goals of my internship included:

- **Understanding the GenCast model**: I studied how it works, what data it uses, and how it compares to traditional forecasting systems.
- Identifying limitations: I evaluated where the model performs well and where it struggles, especially in terms of accuracy, scalability, and reliability.
- **Building a demo**: I developed a prototype that showcases GenCast's capabilities. This demo serves as a proof of concept that can be presented to clients interested in AI-driven forecasting.
- **Client engagement**: The demo is designed to spark interest among clients and help them envision potential use cases. If the technology proves viable, it could lead to future collaborations or product development.
- **Strategic awareness**: Even if GenCast is not yet ready for production use, understanding its current state helps Cognizant stay ahead of technological trends and prepare for future opportunities.

This project allowed me to combine technical exploration with strategic thinking, bridging the gap between cutting-edge AI research and practical business applications.

6. Project Specifications Document

6.1 Project Title

Personalized Weather Forecast Video Generator using GenCast and AWS

6.2 Context : GenCast

GenCast is a cutting-edge probabilistic weather prediction model developed by DeepMind. Unlike traditional numerical weather prediction (NWP) systems, GenCast uses machine learning trained on decades of historical weather data to generate global forecasts. It produces 15-day forecasts in just 8 minutes, with 12-hour intervals and high spatial resolution (0.25° latitude-longitude), covering over 80 atmospheric and surface variables.

GenCast has demonstrated superior performance compared to the world-leading ECMWF ensemble forecast (ENS), outperforming it on 97.4% of evaluated targets. It is particularly effective at predicting extreme weather events, tropical cyclones, and wind power production. By offering faster and more accurate forecasts, GenCast represents a major advancement in operational weather forecasting, enabling better-informed decisions in weather-sensitive sectors.

Source : Deep Minds open source Github repo, including the GenCast research paper and the Opensource model code and demo : https://github.com/google-deepmind/graphcast

6.3 Project Overview

This project, conducted as part of an internship at Cognizant UK, aims to explore and demonstrate the capabilities of DeepMind's GenCast weather prediction model. The goal is to design and develop a prototype application that generates personalized, smartphone-format weather forecast videos based on user-selected dates and locations. These videos will include visual elements such as maps and graphs, natural voice narration, and a virtual avatar to deliver the forecast. The forecast will cover a 72-hour period (the selected day plus two additional days) and will also communicate the uncertainty of the predictions.

The entire solution will be developed and hosted using AWS services, ensuring scalability, performance, and ease of deployment. A key part of the project involve evaluating the performance of different GenCast models, identifying the most suitable one for real-time applications, and exploring methods to increase the temporal resolution of predictions from 12-hour intervals to hourly forecasts. Additionally, the project will investigate the best sources of real-time weather data to feed into the GenCast model, ensuring accurate and up-to-date predictions.

The final deliverable will serve as a showcase for potential clients, demonstrating how GenCast can be adapted to meet specific business needs in sectors where weather forecasting is critical.

6.4 Objectives

The primary objectives of this project are to explore, evaluate, and adapt DeepMind's GenCast weather prediction model for practical, client-facing applications. A key deliverable is the development of a prototype that generates personalized weather forecast videos, tailored to user-selected locations and dates. This prototype will serve as a demonstration tool to showcase the potential of GenCast-powered forecasting solutions to clients across various industries, helping them envision how such technology can be customized to meet their specific operational needs.

6.5 Scope of Work

<u>In Scope:</u>

- Integration with the GenCast model (1deg model).
- Development of a user interface for selecting date and location.
- Generation of 1–2 minute smartphone-format videos with:
    - Forecast for selected day + 2 following days (72 hours).
    - Graphs, maps, and visual elements.
    - Natural voice narration.
    - Display of model uncertainty.
    - Visualise UK-specific output, even if the model runs globally
- Hosting and deployment on AWS.
- Performance evaluation of different GenCast models.
- Exploration of data sources for real-time weather input.
- Investigation into increasing prediction granularity from 12h to 1h.
- Evaluate methods for increasing time resolution.
- Virtual avatar Integration.

<u>Out of Scope:</u>

- Generate forecasts using Google's GenCast 0.25deg model.
- Extend forecasts to 15 days, not relevant for near-term, actionable logistics use cases.
- Rely on metadata manipulation (time-dilation) as a core solution, experimental only.
- Retrain or fine-tune the GenCast model on UK-only input data, global context is required for model validity.
- Develop real-time delivery platforms (e.g. mobile apps or live API services).

6.6 Deliverables

- Functional web-based UI.
- Automated video generation pipeline.
- Evaluation report on GenCast model performance.
- Documentation of AWS architecture and deployment.
- Demo video(s) for client showcase.
- Final presentation and project report.
- Video presentation of the project and its context made by Cognizants Marketing team

6.7 Constraints

The project faces several technical constraints that must be addressed to ensure a smooth and effective implementation. First, the GenCast model currently provides forecasts at a granularity of 12-hour intervals, which may limit the temporal resolution needed for detailed visualizations and user insights. Additionally, the availability and licensing of real-time weather data are critical factors, as the model relies on up-to-date observational inputs to generate accurate forecasts. Another key consideration is the video generation time, which must be optimized to ensure a seamless user experience without long delays. Finally, since GenCast is developed by Google and the project is being built on AWS infrastructure, ensuring compatibility and smooth integration between the two platforms is essential.

6.8. User Case: Forecasting for Offshore Wind Farms

One of the key user cases for this project is enabling offshore wind farm operators to receive personalized weather forecasts and performance insights. The user will be able to select a specific wind farm and a date, and the system will generate a short video forecast covering the selected day and the following two days (72 hours total).

Omar
Renewable Energy Operator
Yorkshire

Goals

- Predict power output to align with commitments
- Schedule maintenance during low-wind periods

Pain Points

- Forecasts are too generic for turbine-level planning
- Missed production estimates cause grid penalties

How GenCast Helps

- Provides micro-local precision at high granularity
- Offers clear visuals to aid decision-making

The video will include:

- Wind conditions (speed, direction, gusts) at the wind farm location
- A performance score estimating how favorable the conditions are for energy production
- Estimated energy output, calculated based on the number and model of wind turbines at the site
- Visualizations such as output curves, wind maps, and uncertainty indicators
- A natural voiceover and virtual avatar presenting the forecast in a smartphone-friendly format

This use case demonstrates how the system can go beyond generic weather forecasting to deliver domain-specific insights that help clients in the renewable energy sector make informed operational decisions.

6.9 Technical Requirements

To successfully complete the project the technical skills required are the following :

1. AWS services (e.g., Lambda, S3, EC2, Bedrock, Sagemaker).
2. Python Integration
3. Python Integration with GenCast model outputs.
4. Frontend framework (e.g., React, Vue).
5. Text-to-speech and avatar generation tools.

6. Data ingestion pipeline for real-time weather data.
7. HTML and CSS for the UI

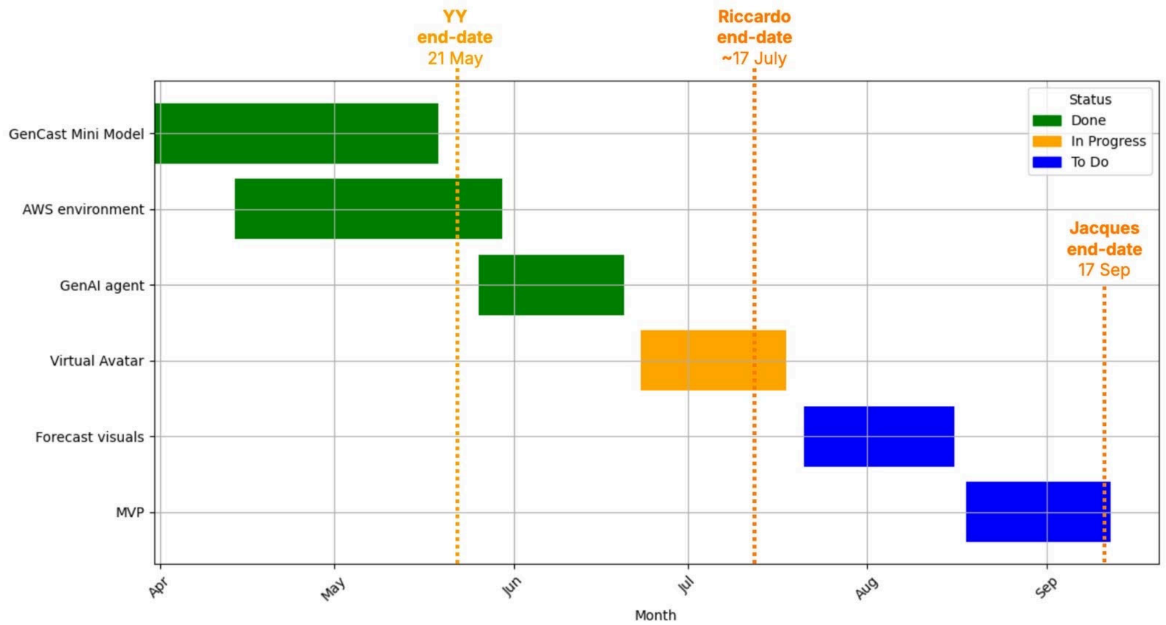
6.10 Success Criteria

To consider the project successful, the system must be capable of generating personalized weather forecast videos within a short response time, ideally under a defined threshold to ensure a smooth user experience. The forecasts produced should maintain a high level of accuracy, staying within acceptable error margins relevant to the use case. Additionally, the solution should receive positive feedback from internal stakeholders and potential clients, demonstrating its value and usability. Finally, the system must be designed with scalability and maintainability in mind, ensuring it can adapt to future enhancements and increased demand.

6.11 Timeline

The project is planned to be completed in 6 months, starting the 17th of April 2025. The project will be divided in milestones, one milestone corresponding to one month of work :

1. Getting the smallest GenCast 1x0deg Mini <2019 model operational in Cognizant environment.
2. Learning GCP cloud environment to operate larger models remotely via BigQuery and VertexAI.
3. Learn how to build GenAI agents. Build a weather forecaster agent to represent the outputs provided by GenCast
4. Learn and create a virtual avatar using our SoulMachines licence to overlay
5. Create visuals of a weather forecast over England.
6. Merge the avatar and visual forecast to represent a demo of the results and capabilities, and limitations.



6.12 Risks and Mitigation

Risk	Description	Mitigation Strategy
Model Accuracy	The GenCast model may not provide forecasts with sufficient accuracy for all use cases, especially in complex or offshore environments. This could reduce the reliability of the forecast videos.	Conduct early-stage evaluation of the model’s performance across different regions and timeframes. If necessary, use fallback models or ensemble methods to improve reliability.
Model Resolution	GenCast currently provides predictions at 12-hour intervals, which may not be granular enough for smooth visualizations or detailed energy output calculations.	Explore interpolation techniques or post-processing methods to simulate hourly forecasts. Investigate whether model fine-tuning or hybrid approaches can improve temporal resolution.
Prediction Variable Compatibility	The model may not output all the variables required for the video content, such as wind speed, direction, precipitation, or cloud cover. This could limit the richness of the forecast.	Identify all required variables during the design phase. If the model lacks certain outputs, supplement with external weather data sources or APIs to fill the gaps.
Real-Time Input Data Availability	The GenCast model relies on real-time observational data, which may be difficult to access, delayed, or subject to licensing restrictions.	Research and secure access to reliable, up-to-date weather data sources early in the project. Prioritize open datasets or establish partnerships with data providers.
AWS Compatibility	Some components of the system, such as avatar rendering or video generation, may not integrate smoothly with AWS services, potentially causing delays or technical issues.	Prioritize the use of AWS-native tools (e.g., MediaConvert, Polly, Lambda). Prototype early to identify integration challenges and adjust architecture accordingly.
Team Changes / Schedule Disruptions	As this is an internship project, changes in team availability or scheduling conflicts could disrupt progress or continuity.	Maintain thorough documentation, use version control (jira), and adopt agile practices to ensure flexibility and knowledge transfer.

7. Presentation and promotion of the work carried out

7.1 Researches

A significant part of my initial responsibilities during the internship involved researching and understanding Google DeepMind’s GenCast model. The objective was to gain a deep understanding of how the model functions in order to later adapt and fine-tune it for our specific use case. Below is a brief overview of the key insights and findings from this research.

7.1.1 GenCast

GenCast is a probabilistic GenAI weather forecasting model developed by DeepMind released on December 4, 2024. It leverages machine learning to produce high-resolution, long-range forecasts with speed and accuracy.

To train GenCast, DeepMind used data from ERA5, a high-quality global climate reanalysis dataset produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) through the Copernicus Climate Change Service (C3S). ERA5 combines decades of weather observations with modern forecasting models to create a detailed, hourly record of the Earth's climate. The GenCast 0.25° <2019 model was trained on ERA5 data from 1979 to 2018 and evaluated on data from 2019 onwards.

GenCast can generate 15-day forecasts in just 8 minutes, with 12-hour intervals and a spatial resolution of 0.25° (latitude-longitude). It covers over 80 atmospheric and surface variables and has demonstrated superior performance compared to the ECMWF ensemble forecast (ENS), outperforming it on 97.4% of evaluated targets.

GenCast is an autoregressive model. It operates by taking two atmospheric snapshots, 12 hours apart, and using them to predict the possible weather scenarios 12 hours into the future. For example, it uses data from time $t-12$ and t to predict scenarios for $t+12$. Then, it uses t and the mean of the predicted $t+12$ scenarios to forecast $t+24$, and so on, until the desired forecast horizon is reached.

The model uses 18 input variables to make predictions, though only 12 of them are predicted. These variables are stored in multidimensional .nc (NetCDF) files, with data available for 13 pressure levels across all geographical points on Earth at a 0.25° resolution.

Input/Output GenCast variable sum up

Type	Variable name	Short name	ECMWF Parameter ID	Role (accumulation period, if applicable)
Atmospheric	Geopotential	z	129	Input/Predicted
Atmospheric	Specific humidity	q	133	Input/Predicted
Atmospheric	Temperature	t	130	Input/Predicted
Atmospheric	U component of wind	u	131	Input/Predicted
Atmospheric	V component of wind	v	132	Input/Predicted
Atmospheric	Vertical velocity	w	135	Input/Predicted
Single	2 metre temperature	2t	167	Input/Predicted
Single	10 metre u wind component	10u	165	Input/Predicted
Single	10 metre v wind component	10v	166	Input/Predicted
Single	Mean sea level pressure	msl	151	Input/Predicted
Single	Sea Surface Temperature	sst	34	Input/Predicted
Single	Total precipitation	tp	228	Predicted (12h)
Static	Geopotential at surface	z	129	Input
Static	Land-sea mask	lsm	172	Input
Static	Latitude	n/a	n/a	Input
Static	Longitude	n/a	n/a	Input
Clock	Local time of day	n/a	n/a	Input
Clock	Elapsed year progress	n/a	n/a	Input

For a single timestamp, the data volume is:

= 721 latitude points × 1440 longitude points × 13 pressure levels × 18 variables
= 243,164,160 values

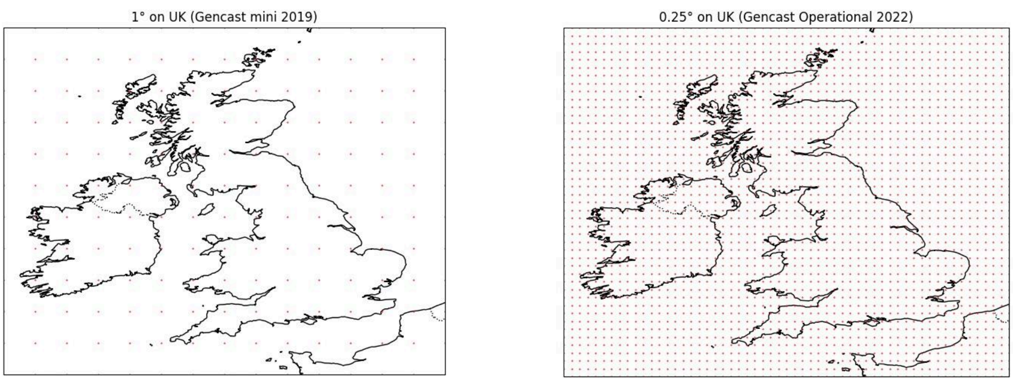
A full 15-day forecast (30 time steps at 12-hour intervals) can reach up to 11.2 GB of data.

Due to the scale and complexity of the data and computations, GenCast relies on TPUs (Tensor Processing Units)—specialized processors developed by Google for accelerating machine learning workloads, particularly tensor operations in neural networks and scientific computing. However, TPUs are resource-intensive and not ideal for daily use in all environments. To address this, DeepMind developed lighter variants of the model:

- GenCast 1.0° <2019
- GenCast 1.0° Mini <2019

These versions have a lower spatial resolution (180×360) and a smaller memory footprint, making them more accessible for practical use, albeit with reduced geographical precision.

Resolution difference between the 1x0 model and 0x25 :



7.1.2 Data Input and Performance

Data Integration and Real-Time Forecasting :

As mentioned earlier, GenCast requires input data to generate forecasts. This means that to perform real-time predictions—such as forecasting three days ahead— it has to access a data source that provides atmospheric snapshots no older than 12 hours. After completing the initial research on the model, my next task was to identify a suitable data source, ideally an API, that could be directly integrated with the model.

The goal was to enable our final product to support two modes:

- Historical Evaluation: The user could input a past date and a target forecast date to simulate predictions using historical data. This would allow us to compare the model's output with actual weather data and evaluate its accuracy.
- Real-Time Forecasting: The user could select the current date to generate live forecasts.

In both cases, the system would automatically fetch the required data from the source and feed it into the model. However, this posed a significant challenge: GenCast requires input data in a very specific format, making it difficult to find a compatible data source.

While DeepMind provides a small set of preprocessed datasets on their public GitHub repository to demonstrate GenCast, these only cover a limited 30-day period starting from March 29, 2019. This was insufficient for our needs, as we wanted the flexibility to run the model from any date.

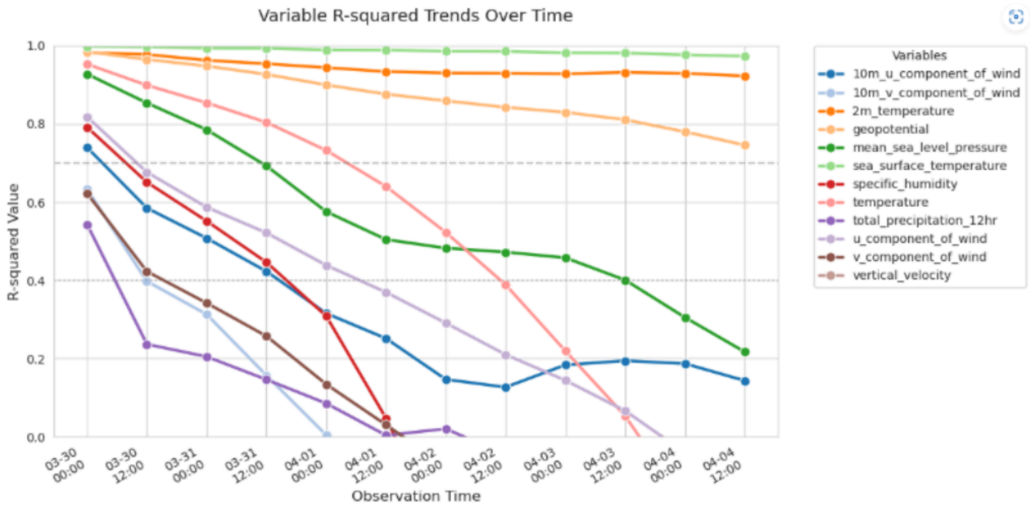
To address this, we established a connection to the Copernicus Climate Data Store (CDS) API to fetch raw ERA5 data—the same dataset used to train GenCast. However, this raw data does not match the WeatherBench2 format expected by the model. As a result, we had to invest significant effort into reformatting and preprocessing the data to make it compatible.

Unexpected Results and Model Performance :

Once we successfully integrated our custom-prepared data with GenCast, we ran a series of tests comparing the model's forecasts to actual weather data. Surprisingly, the results were underwhelming. Despite GenCast's reported superiority over traditional ensemble forecasts (ENS) beyond 36 hours, our experiments showed R^2 scores dropping below 0.4 for some variables after just three days.

This discrepancy raised important questions about the model's generalizability and sensitivity to input data quality and formatting. It also highlighted the challenges of replicating state-of-the-art research models in real-world applications, especially when working with limited documentation and strict data requirements.

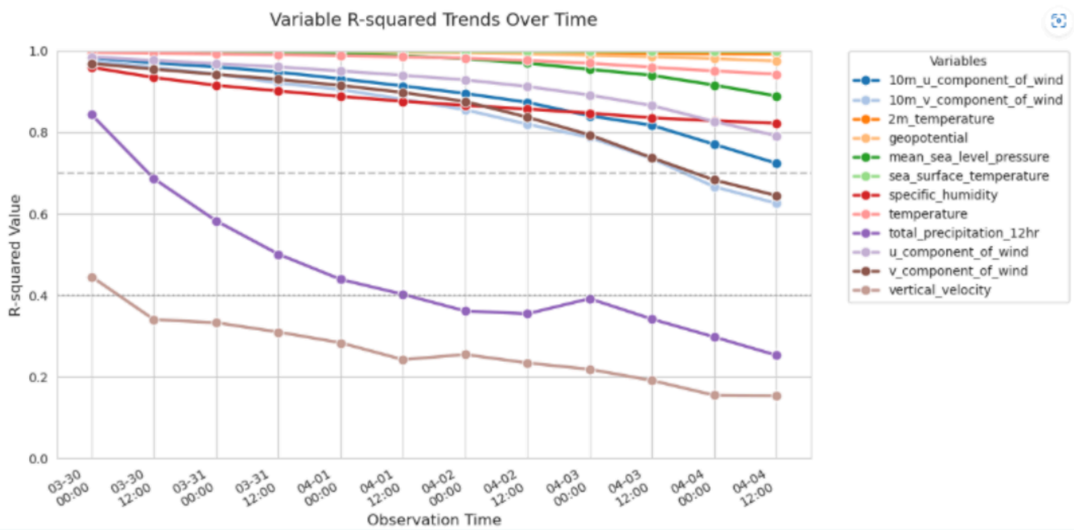
R-square performance metric of GenCast prediction using CDS API data



However, when we ran the model using DeepMind's official preprocessed data for the same dates, the model performed exactly as expected, closely matching the results reported in their paper. This strongly suggests that the issue lies not with the model itself, but with the quality or formatting of our input data.

This experience highlighted the critical importance of data preprocessing in machine learning pipelines—especially for complex models like GenCast. Even small inconsistencies in data structure, resolution, or variable alignment can significantly degrade performance.

R-square metric of GenCast prediction using DeepMind's official preprocessed data



Resolving Data Quality Issues and Improving Model Performance :

After encountering unexpectedly poor performance when running GenCast with our custom-prepared ERA5 data, we reached out to the DeepMind team through a developer forum. Through these discussions, we identified that the issue likely stemmed from how we were regridding the input data.

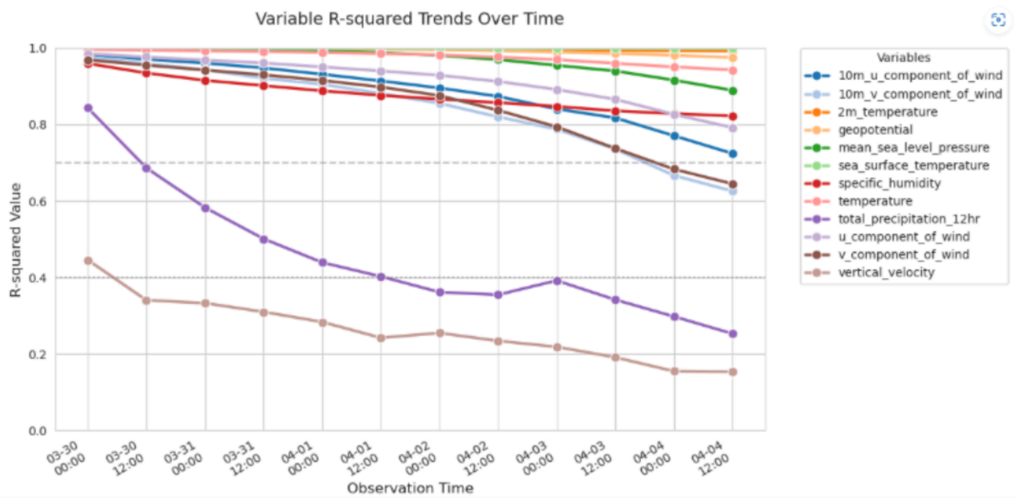
GenCast is available in multiple spatial resolutions, including GenCast 1.0° and GenCast 0.25° (for data prior to 2019). While both versions use the same set of variables, they differ significantly in geographical resolution. The 0.25° model makes predictions at four times as many grid points as the 1.0° model, making it much more demanding in terms of compute and memory.

Importantly, the GenCast models were trained on 0.25° resolution data. The 1.0° datasets are derived by subsampling the 0.25° data. However, depending on how this subsampling is performed, subtle variations can be introduced that negatively impact model performance.

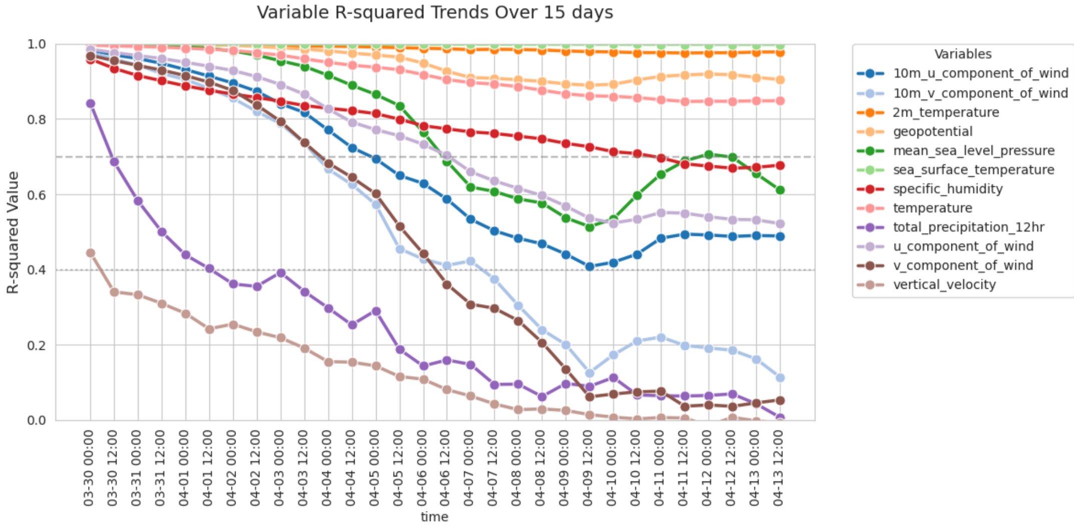
To reduce the risk of preprocessing errors, DeepMind recommended using a Google Cloud Platform (GCP) bucket that hosts ERA5 data already formatted to match the WeatherBench2 standard at 0.25° resolution. This approach minimizes the chances of misalignment or formatting issues and ensures compatibility with the model's expectations.

By switching to this new data source and handling the subsampling ourselves, we were able to launch GenCast from any starting date up to 6 days behind real-time, achieving strong performance across multiple 15-day forecast periods. This significantly improved the reliability and flexibility of our experiments.

3 days prediction with input from the bucket recommended by the DeepMind team :



15 days prediction with input from the bucket recommended by the DeepMind team :



Variable-Specific Performance:

As shown in the graphs above, GenCast's predictive performance varies depending on the specific atmospheric variable. Some variables, such as temperature and pressure, tend to be more stable and easier to forecast, while others—like wind speed or precipitation—are inherently more volatile and challenging to predict accurately.

Despite these differences, the overall results are very promising. The model consistently demonstrates strong performance across most variables, especially when using high-quality, properly formatted input data. This reinforces the robustness of GenCast and its potential for reliable medium-range weather forecasting.

7.1.3 Time resolution

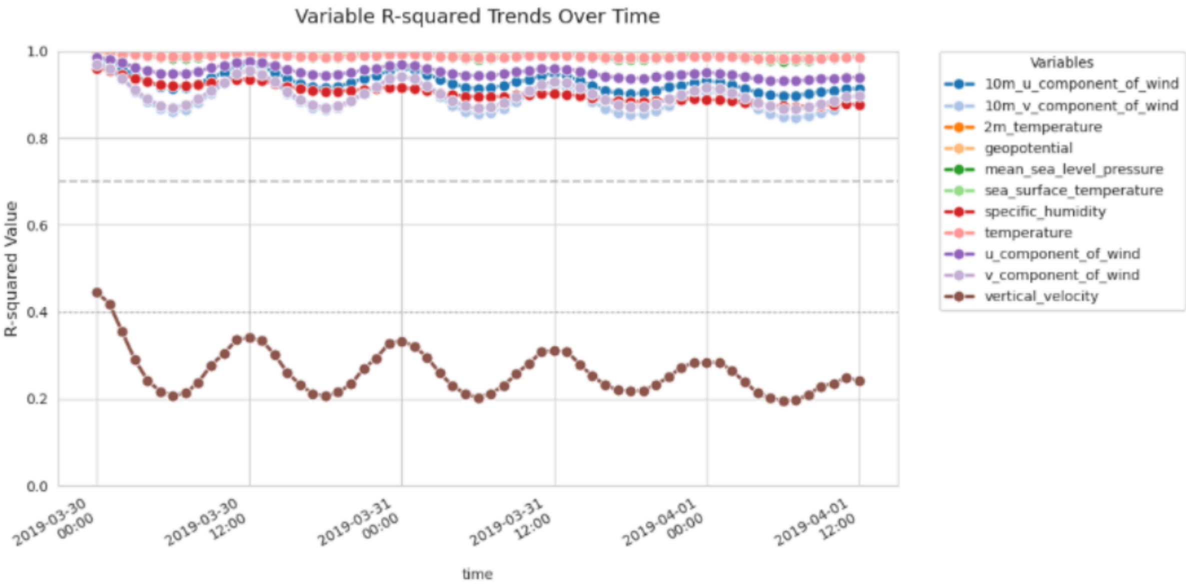
The GenCast model operates with a time resolution of 12 hours, meaning it produces weather snapshots at three fixed times each day: midnight (00:00), midday (12:00), and midnight again (24:00). While this setup is sufficient for many forecasting applications, it posed a significant limitation for our use case. Over a three-day forecast period, we would only obtain seven data points, which is far too sparse for generating smooth, continuous visualizations or for analyzing the detailed evolution of energy production throughout the day. This lack of temporal granularity made it difficult to create engaging, informative outputs such as animated weather maps or intra-day energy forecasts.

To address this, we initially explored the idea of launching multiple GenCast models with staggered start times. For example, running two models offset by six hours would double the temporal resolution, and running four would bring it down to three-hour intervals. We could then apply linear interpolation to reach an hourly resolution. While this approach showed promising results in terms of accuracy, it significantly increased the computational load, storage requirements, and processing time. From a user experience perspective, this was not acceptable, as each forecast would take longer to generate and exceed our target of delivering results in under ten minutes.

Another option we considered was training a custom machine learning model to interpolate between GenCast outputs and generate hourly predictions. This approach had theoretical appeal, especially given the large volume of training data available. However, it would have required several weeks of development, substantial GPU resources, and a complex data pipeline. Moreover, the performance of such a model would remain uncertain until fully implemented, making it a high-risk investment within the constraints of our project timeline.

Ultimately, we chose to use a simpler and more efficient method: direct interpolation of the 12-hour GenCast outputs. We tested several interpolation techniques, including linear, quadratic, and cubic methods. After evaluating their performance, we selected linear interpolation as the most suitable option. It offered fast computation, stable and predictable results, and avoided the risk of introducing artifacts or overfitting. This solution provided the best balance between performance, simplicity, and responsiveness, allowing us to generate smooth, hourly-resolution forecasts without compromising the user experience.

R square of linear interpolation over 3 days :



7.1.4 AWS Compatibility and inference choice :

Although GenCast is a model developed by Google DeepMind and optimized to run on Google Cloud Platform (GCP) using TPUs, we chose to host the project on Amazon Web Services (AWS). This decision was primarily driven by familiarity and prior experience. At the start of the project, we were unaware of the performance differences between GCP and AWS for this specific use case. I had already completed two AWS Quest certifications and invested significant time learning the platform, so switching to GCP would have meant starting from scratch with a new cloud environment. Given the tight timeline and the need to move quickly, continuing with AWS was the most practical choice.

GenCast is built using JAX, a high-performance numerical computing library similar to PyTorch, which enables parallelized computation. However, the model was specifically designed to run on TPUs, which are only available on GCP. AWS does not currently offer TPU equivalents, which introduced a challenge in finding suitable infrastructure to run the 0.25° resolution version of GenCast.

According to DeepMind’s documentation, inference on TPUs using the optimized splash_attention mechanism takes approximately 8 minutes for a 30-step rollout. When using the alternative triblockdiag_mha attention mechanism, inference time increases to 15 minutes on TPUs and 25 minutes on GPUs. Since AWS only supports GPU-based inference, we were limited to the slower configuration.

In addition to longer inference times, GPU-based inference also requires significantly more memory. Running GenCast 0.25° on a GPU demands approximately 300 GB of system memory and 60 GB of VRAM, while the 1.0° version requires only 24 GB of system memory and 16 GB of VRAM. This made instance selection critical.

We tested several AWS SageMaker instances that theoretically met the memory requirements. However, many failed during runtime due to memory allocation errors. For example, the g5.24xlarge and ml.p3.16xlarge instances both returned XlaRuntimeError: RESOURCE_EXHAUSTED, despite having sufficient memory on paper. The g5.48xlarge and ml.p4de.24xlarge instances were unavailable in our region (London AZ), and attempts to use multi-GPU setups revealed that memory was not being distributed efficiently across GPUs.

Despite meeting the technical specifications, we were unable to successfully launch the 0.25° model on AWS. This left us with three possible paths forward. The first was to continue troubleshooting the 0.25° model on AWS, which would be time-consuming and expensive, especially since we planned to run multiple GenCast instances to achieve higher temporal resolution. The second option was to switch to the 1.0° model, which runs significantly faster (around 4 minutes per prediction), is less resource-intensive, and still provides visually acceptable results—particularly useful for storytelling and demonstration purposes. The third option was to migrate the project to GCP to take advantage of TPUs and optimized inference, but this would require a steep learning curve and a longer development timeline.

We ultimately chose to proceed with the 1.0° model on AWS, as it offered the best balance between performance, cost, and development speed. Importantly, this decision remains flexible. If needed, we can revisit the 0.25° model or migrate to GCP later in the project without major disruption.

To further validate our choice, we ran 12-step (6-day) predictions on various AWS instances. The ml.g5.12xlarge instance provided the best performance-to-cost ratio, completing predictions in under 9 minutes at a reasonable hourly rate. Other instances, such as ml.p3.2xlarge, were slower and more expensive, while smaller instances like c8g.4xlarge failed to meet the time constraints required for a responsive user experience.

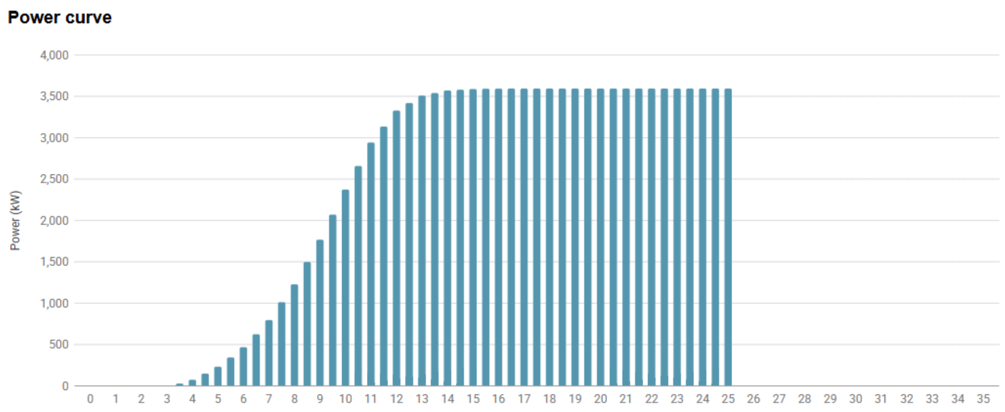
7.1.5 Wind farm energy production :

To ensure the information presented in our forecasts was truly relevant to our use case, we needed to go beyond the standard weather variables typically found in conventional forecasts—such as temperature, wind, and rainfall. Our goal was to identify which environmental factors actually impact wind turbine performance, and tailor our visualizations and metrics accordingly.

Through research and analysis of wind turbine documentation, we found that wind speed and direction are the only variables that directly influence turbine output. Other factors like temperature, pressure, or rainfall do not significantly affect performance. Based on this insight, we focused our efforts on modeling and visualizing wind behavior in detail.

Using the turbine's power curve, which maps wind speed to energy output, we were able to determine the optimal operating conditions and calculate the expected power generation for each forecasted time step.

Power Curve of a SWT-3.6 -107 Wind Turbine depending on the Weed Spend



Srouce : https://www.thewindpower.net/turbine_en_20_siemens_swt-3.6-107.php

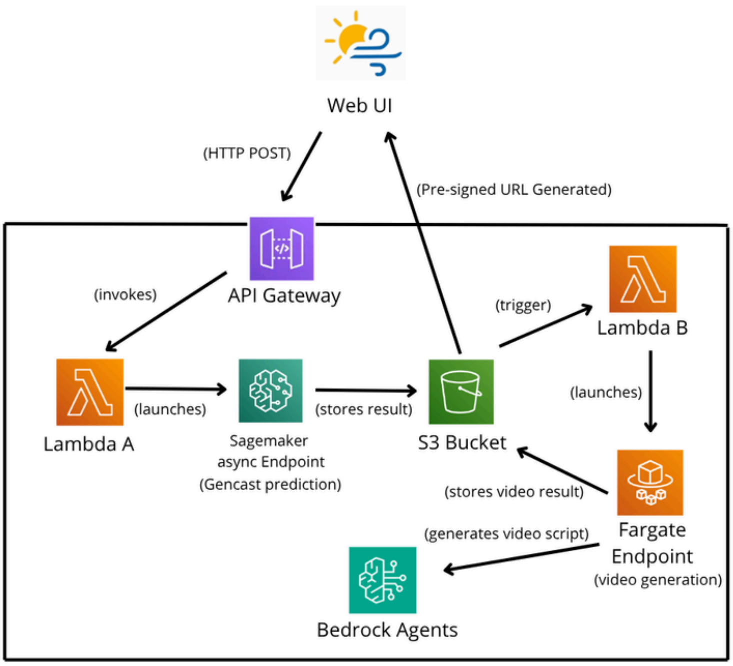
Although not directly related to turbine output, we decided to include added visuals showing rainfall, estimating these would also be of interest to the user in helping to predict dry periods when turbine maintenance might be possible.

This allowed us to provide not just weather data, but meaningful performance insights that are directly useful for wind farm technicians and planners.

7.2 Presentation of the implemented solution

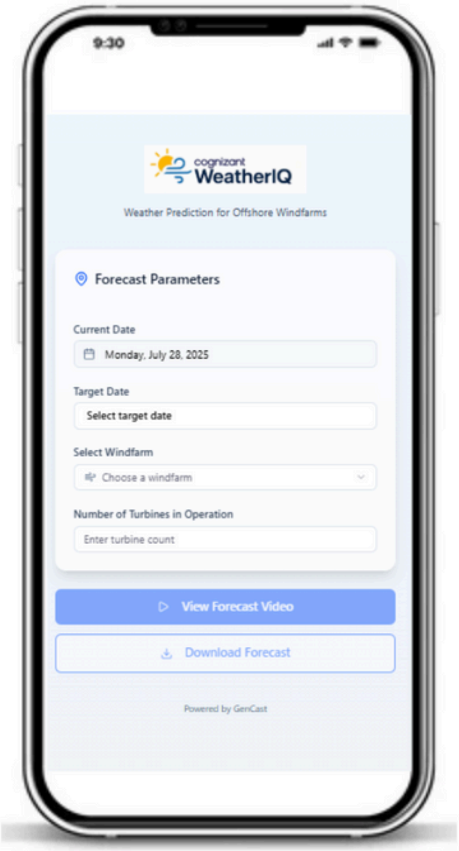
7.2.1 AWS architecture

The system is built on a serverless, event-driven architecture using Amazon Web Services (AWS), ensuring scalability and cost-efficiency. Users interact with a web-based interface to input key parameters such as the current date, a target forecast date, the selected wind farm, and the number of turbines in operation. Once submitted, this data is processed by an AWS Lambda function, which packages the input and sends it to an Amazon SageMaker asynchronous inference endpoint hosting the GenCast model. This model generates a 3-day weather forecast tailored to the selected wind farm. The output is stored in an Amazon S3 bucket. The arrival of this data triggers an S3 event, which invokes a second Lambda function responsible for orchestrating the next stage: launching a containerized video generation task on AWS Fargate via Amazon ECS. This task runs a Python application that reads the forecast data, generates a video visualizing the weather and estimated energy production with the help of Bedrock agents, and stores the final output back in S3. The video is then made accessible to the user through a secure pre-signed URL.



7.2.2 Current Project Status

As I am writing this report before the end of my internship, the complete system architecture is not yet fully operational. However, all the individual components are in place and ready for use. The code has been containerized, the user interface is functional, and the necessary AWS Lambda functions have been created. What remains is the integration work—linking all these components together into a seamless, automated pipeline. This final step requires some additional development and testing to ensure smooth communication between services and a reliable end-to-end experience. The documentation has been structured to ensure continuity, enabling a smooth handover to future interns or project teams who may wish to build upon this foundation.



7.2.3 Forecast video :

7.2.2.1 Automated Video Generation Pipeline:

The final stage of the project involved building a fully automated pipeline to generate dynamic weather forecast videos tailored to a specific wind farm location in the UK. The process began with receiving and processing UK weather data, which included cleaning the raw inputs and calculating derived variables such as total rainfall, wind direction (both in degrees and cardinal format, derived from u and v wind components), and turbine-specific metrics like energy output and performance percentage.

Once the data was cleaned and enriched, we moved on to data sampling. This involved selecting relevant time windows—typically three-day periods—and extracting key variables related to energy production and weather conditions. These samples were stored in JSON format and served as input for generative AI agents.

We then developed a set of Bedrock agents, each responsible for generating the script for a specific section of the video. These sections included the introduction, morning and afternoon forecasts, a three-day overview, an energy-focused segment, and the outro. All scripts were generated using Amazon’s “Nova Pro” model, which provided high-quality, context-aware text generation.

For the audio narration, we used Amazon Polly, a text-to-speech service, with the “Amy” voice model to convert the generated scripts into natural-sounding speech. This allowed us to maintain a consistent and professional tone throughout the video.

The visual component of the video is composed of several elements. First, we generated dynamic weather maps showing the evolution of rainfall, wind, and turbine performance across the UK. These maps were animated to reflect changes over time. In parallel, we created a dynamic table using the Manim library to display three-day forecasts in a clear and engaging format.

Finally, all components were brought together in the global video assembly stage. This involved synchronizing the audio narration with the visual elements and concatenating all video segments into a seamless final product. The result was a fully automated, data-driven video that could be generated on demand for any selected date and location.

7.2.2.1 Automated Video Result :

The final video is composed of several distinct subsections, each designed to convey specific information in a clear and engaging way.

The introduction serves as a brief opening sequence, lasting only a few seconds. It features an animated version of the Cognizant WeatherIQ logo, establishing the visual identity of the product. Simultaneously, a voiceover introduces the selected forecast date and offshore wind farm location. It also provides a concise summary of the expected weather conditions over the wind farm, setting the stage for the detailed forecast that follows.

"Here is your Cognizant weather IQ forecast for April 1 2019 at the Race Bank windfarm. Expect a day with moderate winds and varying turbine performance levels."

The second part of the video presents a detailed weather forecast in a style inspired by traditional television broadcasts. The narration guides the viewer through the day in two parts—morning (AM) and afternoon (PM)—providing a clear and structured overview of the evolving weather conditions. For each part of the day, the video displays dynamic visualizations illustrating rainfall, wind patterns, and turbine performance over time.

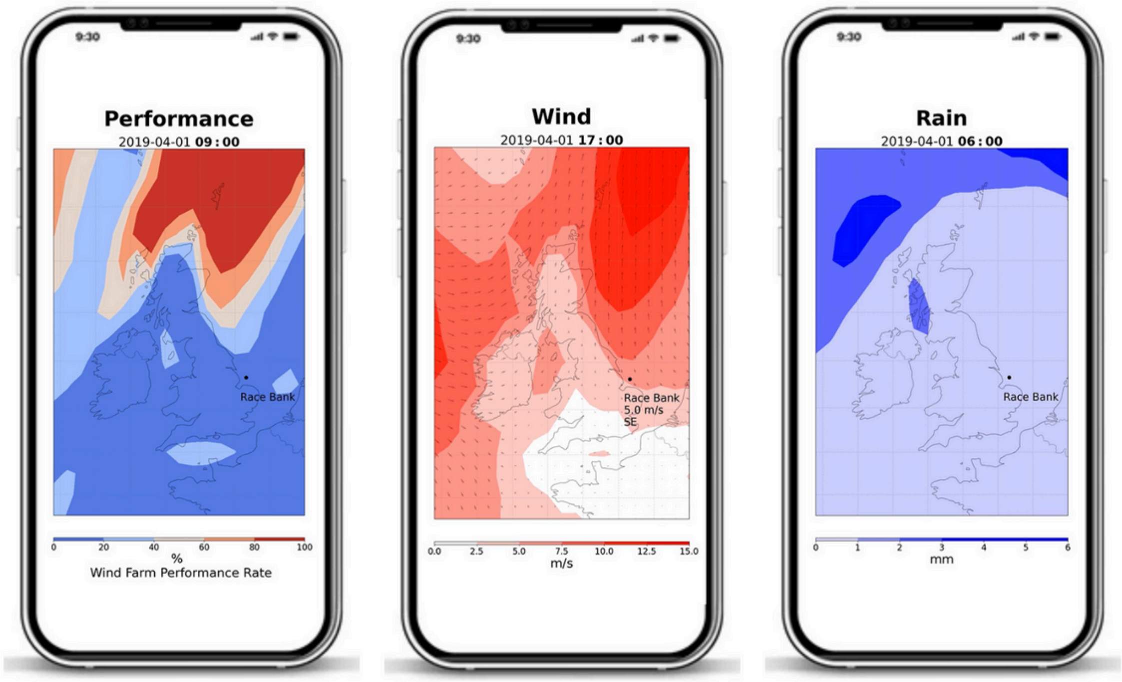
One of the main challenges in this section was ensuring that the voiceover remained synchronized with the visuals, maintaining a smooth and coherent flow throughout the segment. This required careful timing and alignment between the generated audio and the animated weather data. Visually, the forecast is brought to life through dynamic transitions that reflect the passage of time, allowing viewers to intuitively follow the progression of weather conditions across the day.

We made a deliberate choice not to include a map of temperature evolution in this section, as this variable has minimal impact on turbine performance and would not add significant value to the viewer. Instead, we prioritized variables like rainfall, which can directly affect maintenance planning and operational logistics. This decision helped keep the video concise and focused on the most relevant and actionable information.

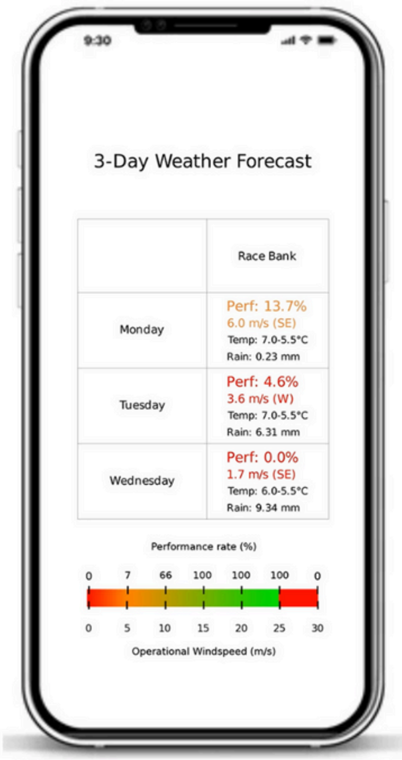
The performance map displays the geographical areas surrounding the wind farm along with the estimated turbine performance rates in each zone. These rates are calculated based on the local wind conditions and the turbine's power curve, allowing us to visualize how efficiently the turbines would operate if placed in those specific areas.

"Starting with the morning, we see no rain expected before midday, with total precipitation remaining at 0.0 mm. Winds will be coming from the west at 5.7 m/s, shifting to the northwest at 5.6 m/s by midday. Turbine performance will start at 10.9% and slightly dip to 10.6% by midday.",

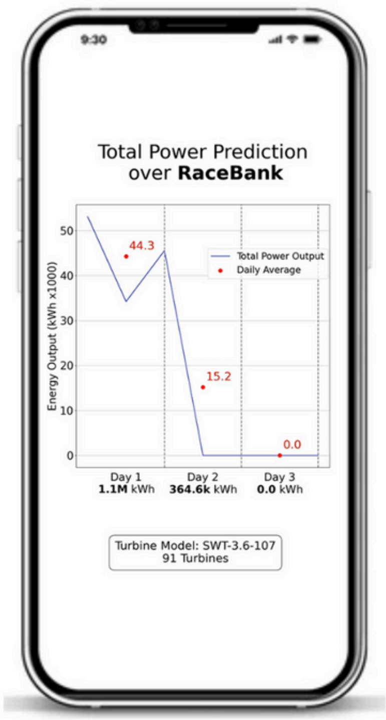
"Moving into the afternoon and evening, we anticipate a slight increase in precipitation, with 0.2 mm expected by 5 PM and continuing through midnight. Wind speeds will decrease to 5.0 m/s from the northwest, then shift to the north at 5.3 m/s by 9 PM, increasing to 6.1 m/s by midnight. Turbine performance will drop to 6.7% by 5 PM, improve to 9.0% by 9 PM, and significantly increase to 13.9% by midnight."



Following the detailed forecast of the selected day, the video transitions into a broader overview of the weather conditions for the two subsequent days. This section is presented in a more concise format, with the information displayed in a tabular layout that summarizes key variables such as rainfall, wind, temperature and turbine performance. While the visuals remain static and structured, the narration continues to guide the viewer through the data, highlighting notable trends and changes. This approach allows for a quick yet informative glance at the upcoming days, maintaining the flow of the video while preparing the viewer for the final segments.



"Over the next 3 days at Race Bank, expect a slight drop in temperatures and varying conditions. On Monday, it will be cool with a high of 7\u00b0C and a low of 5.5\u00b0C, very light rain of 0.23 mm, and moderate winds from the west at 6 m/s, with wind turbines at 13.7% performance. Tuesday will see similar temperatures with a high of 7\u00b0C and a low of 5.5\u00b0C, light rain of 6.31 mm, and lighter winds from the south at 3.6 m/s, reducing turbine performance to 4.6%. By Wednesday, temperatures will slightly drop to a high of 6\u00b0C and a low of 5.5\u00b0C, with moderate rain of 9.34 mm and very light winds from the southwest at 1.7 m/s, leading to a complete shutdown of wind turbines at 0% performance.\n"



The final section of the video focuses on a deeper analysis of energy production forecasts over the three-day period. This segment presents a curve showing the total energy output per hour, allowing viewers to observe fluctuations in production throughout each day. In addition to the hourly breakdown, the video also displays the cumulative energy produced in kilowatt-hours (kWh), offering a clear summary of overall performance.

To provide further context, the video includes key operational details such as the number of active wind turbines and the specific turbine model used at the selected wind farm. These elements help ground the forecast in real-world parameters and give viewers a more complete understanding of the factors influencing energy output.

"Over the next 3 days at the Race Bank offshore windfarm, which consists of 91 turbines, we expect a significant variation in energy production. On Monday, the windfarm is predicted to produce an average of 44.3 kilowatts per hour, totaling 1.1 million kilowatts for the day. However, as we move to Tuesday, there will be a noticeable drop in production, with an average of 15.2 kilowatts per hour and a total of 364.6 thousand kilowatts. By Wednesday, the forecast indicates a complete halt in energy production, with both the average hourly output and the total daily output expected to be zero kilowatts."

The video concludes with a short outro segment that wraps up the forecast and reinforces the identity of the project. The narration briefly summarizes the overall weather situation covered in the video and reminds the viewer of the project name, bringing the experience to a clear and cohesive close. This final touch helps maintain continuity and leaves the user with a strong, professional impression of the WeatherIQ platform.

"Overall, it's a day of moderate winds with a notable shift in direction and varying turbine performance, influenced by the wind conditions. This was your Cognizant weather IQ forecast."

7.3 Difficulties encountered

- Endpoint :

One of the most challenging aspects of the project was deploying the GenCast model on an Amazon SageMaker endpoint. The goal was to host the model on AWS and make it accessible via Lambda functions for inference requests. This required containerizing the model, which involved setting up a custom environment with all necessary dependencies pre-installed.

The most difficult part was designing the interface between the inference logic inside the container and the external environment. Since containers are isolated, they need a defined interface—typically a web server listening on a specific port—to receive and respond to inference requests. SageMaker provides a prebuilt solution for this called TorchServe, which is designed to simplify model deployment. However, despite spending several weeks trying to configure it, I was unable to get TorchServe to correctly launch my custom inference code within the container.

Testing was also complicated by the need for GPU-enabled instances, which are significantly more expensive. The pressure to deliver quickly led to mistakes and inefficiencies, ultimately increasing both time and cost.
Eventually, I pivoted to a more pragmatic approach. I created a lightweight dummy version of the model to test the container logic without requiring GPU resources. This drastically reduced testing costs. I then implemented a custom interface using Flask, which allowed me to control the request handling logic directly. This solution worked reliably, and I was able to save the model outputs to an S3 bucket as intended.

This experience taught me the importance of balancing the use of existing tools with custom development. In hindsight, I spent too long trying to make a plug-and-play framework work, rather than building a minimal custom solution from the start. While frameworks like TorchServe can be powerful, they also introduce complexity and assumptions that may not align with specific project needs.

- Autoregression:

Another significant challenge I faced was implementing autoregression with the GenCast model. Autoregression involves feeding the output of a prediction back into the model as input for the next time step. In the demo provided by DeepMind, the model performs a single 12-hour forecast, and at first, I assumed that the autoregressive logic was not publicly available.

As a result, I attempted to implement it myself. This turned out to be quite complex, primarily because the model's output format did not exactly match the expected input format. Some variables used during inference are not present in the output, and GenCast is particularly strict about input structure and variable consistency. I spent considerable time reverse-engineering the output and adapting it to fit the model's input requirements.

Eventually, I managed to recreate a working autoregression loop. However, I later discovered that I had misunderstood the demo: DeepMind's code already included the autoregressive logic—I had simply been using it incorrectly. I ended up reverting to their implementation, which was more robust and reliable than my custom version.

This misstep cost me several days of work and unnecessary GPU usage, which added to the project's cost. However, it also gave me a much deeper understanding of how the model operates internally and how to handle strict input/output constraints in machine learning pipelines.

- Working in a minimalist team :

A final challenge I encountered was related to the learning environment itself. Cognizant is a large organization, and understandably, most team members are engaged in client-facing projects with tight deadlines. As a result, I often found myself working alone on the technical aspects of my project.

When I ran into complex issues—especially those involving infrastructure, deployment, or debugging—it was sometimes difficult to find someone with the time and availability to provide in-depth support. Consequently I had to troubleshoot and learn many things on my own, which slowed down my progress at times.

While this autonomy helped me grow technically and become more resourceful, it also led to moments of frustration and inefficiency. Hours or even days solving problems could have been resolved much faster with more mentorship or peer review. In the end, this experience taught me the value of asking for help early, but also to persevere through technical uncertainty and strengthened my ability to learn independently.

7.4 Prior knowledge used and skills acquired.

Skill acquired	Application in the Project
AWS Services (Lambda, S3, EC2, Bedrock, SageMaker)	Used throughout the infrastructure setup, model hosting, and agent deployment. SageMaker was specifically used to test and run GenCast models. Bedrock was used for generative agents and audio synthesis.
AWS Certifications (Cloud Practitioner & ML Quest)	Provided foundational knowledge for navigating AWS services efficiently and making informed architectural decisions.
Python Integration	Used for data preprocessing, model input formatting, and pipeline automation. Also used to interface with GenCast outputs and interpolate results.
Python Integration with GenCast	Enabled the execution of GenCast forecasts, handling model inputs and outputs, and managing inference logic.
Frontend Frameworks (React, Vue), HTML/CSS	Applied in the development of the user interface for selecting dates, wind farm locations, and viewing generated videos.
MLOps, Docker, Endpoint Creation	Used to containerize and deploy the model and agents, ensuring reproducibility and scalability. Endpoints were created for triggering forecasts and video generation.
UI/UX Understanding	Guided decisions on what information to display (e.g., excluding temperature maps), ensuring clarity and relevance for end users.
Project Management & Communication	Participated in weekly meetings, presented milestones, and maintained clear documentation of progress, risk mitigation and decisions in an Agile manner.
Text-to-Speech & Avatar Generation Tools	Used Amazon Polly to generate voiceovers for each video segment, maintaining consistency and quality.

Data Ingestion Pipeline for Real-Time Weather Data	Built to fetch, clean, and format ERA5 data for GenCast input, including wind, rain, and turbine performance metrics.
Project Management with Jira	Used Jira to plan sprints and MVPs ahead, coordinate with the project manager and keep track of the timeline.

During my internship, I had the opportunity to deepen my understanding of AWS by completing two AWS Cloud Quest certifications: Cloud Practitioner and Machine Learning. While I had already been introduced to AWS concepts during my academic coursework, these certifications allowed me to apply that knowledge in a hands-on, practical way. AWS Cloud Quest is an interactive learning platform that teaches cloud skills through real-world scenarios and guided challenges. The Cloud Practitioner track helped me build a strong foundation in cloud services and architecture, while the Machine Learning track focused on deploying and managing ML models using tools like SageMaker, Lambda, and EC2.

These certifications were directly relevant to the work I was doing during the internship. I was able to demonstrate and apply the skills I was learning almost immediately, which made the experience both efficient and rewarding. I'm really thankful to the company for supporting and sponsoring me through these certifications—it was a great opportunity to grow professionally and contribute more confidently to the project.

8. Critical assessment and Contribution to the company

8.1 Critical assessment

My internship project contributed to Cognizant’s exploration of AI-driven weather forecasting by evaluating the capabilities and limitations of DeepMind’s GenCast model in a real-world, production-like setting. While the final solution is not yet ready for operational deployment, the work conducted serves as a valuable proof of concept and a technical deep dive into a rapidly evolving field.

Critically, the GenCast model’s spatial and temporal resolution posed limitations for our use case. The 1.0° version used in production lacked the precision needed for localized wind farm forecasting, and the 12-hour timestep between predictions was insufficient for detailed energy output estimation. Additionally, some atmospheric variables—such as cloud formation—were missing, which limits the model’s applicability for broader weather analysis and other user cases.

These limitations reflect broader concerns in the field. AI models, while promising, often underperform at sub-1000 km scales, making them less reliable for detecting localized weather events like troughs, ridges, or flash floods. Their resolution—typically around 28 km—can miss small but impactful phenomena, and their reliance on historical data makes them vulnerable to climate drift and rare event blind spots. Moreover, AI models often lack physical interpretability, making it difficult to understand or trust their predictions in high-stakes scenarios.

However, despite these challenges, AI models remain a valuable complement to traditional forecasting systems. Rather than replacing physics-based models, they can enhance forecasting workflows by providing faster, more scalable predictions—especially for large-scale atmospheric patterns. The most promising path forward is a hybrid approach, where AI and traditional models work together, leveraging their respective strengths to deliver more accurate, timely, and localized forecasts.

8.2 Contribution to the company

Ultimately, while the solution developed during this internship is not yet production-ready, it stands as a well-documented research initiative. The project has been added to Cognizant’s internal portfolio and can be leveraged in future client engagements. Furthermore, the documentation has been structured to ensure continuity, enabling a smooth handover to future interns or project teams who may wish to build upon this foundation—saving time, reducing risk, and accelerating delivery.

Importantly, Cognizant has already identified a promising continuation of this project in collaboration with the UK Met Office. The initiative aims to develop a software solution capable of predicting the emission and dispersion of toxic chemicals in the atmosphere and soil. This has significant implications for public safety and environmental protection—particularly in scenarios such as industrial accidents or chemical lab fires, where hazardous substances could pose serious risks to nearby populations and ecosystems.

The Met Office has an existing atmospheric dispersion model known as NAME (Numerical Atmospheric-dispersion Modelling Environment), which can simulate the spread of chemical pollutants. By integrating this model with the weather forecasting capabilities explored during the internship, Cognizant is well-positioned to contribute meaningfully to the project.

source : https://www.metoffice.gov.uk/research/approach/modelling-systems/dispersion-model

9. Impact on my personal project :

9.1 Understanding the Engineering Profession

The assignment I undertook during my internship closely aligns with what I envision the role of an AI programming engineer to be. It combined several key aspects of the profession: applying technical skills, independently researching unfamiliar domains, and rapidly prototyping solutions to complex problems. I had the opportunity to follow a Project Specifications Document, navigate around constraints, and enjoy the freedom to think creatively and test different approaches—an experience that mirrors real-world engineering challenges.

Although the client interactions in my case were fictional, they provided a valuable simulation of how communication with stakeholders might unfold in a professional setting. This helped me better understand the importance of translating technical work into business value.

One element that was somewhat missing from the full engineering experience was team collaboration. As I was the sole technical contributor on the project, I didn’t get to engage in peer code reviews, collaborative problem-solving, or agile team dynamics. Nonetheless, this experience gave me a strong foundation in independent work and reinforced my interest in pursuing a career in AI engineering.

Additionally, as this was my first time working in a large company, I gained valuable exposure to the broader dynamics of professional life in a corporate environment and a major city. Beyond the technical work, I observed how people interact across departments, how communication styles vary depending on roles, and how informal networks often play a key role in navigating the workplace. I began to understand the different motivations that drive individuals—whether it's a desire for recognition, career advancement, or alignment with company values—and how these motivations influence behavior and decision-making.

I also became more aware of the performance-driven culture that characterizes consulting firms. Success is often measured through metrics such as billable hours, client satisfaction, and project delivery speed. This creates a fast-paced, high-pressure environment where adaptability and responsiveness are essential. While this can be energizing, it also introduces a level of stress that is quite different from what I experienced during my previous internship in the technical department of the army.

In my previous internship, motivation was often rooted in a deep passion for the field and a sense of duty. In contrast, the corporate world felt more dynamic and competitive, with a stronger emphasis on visibility, networking, and strategic positioning within the company hierarchy. This contrast helped me appreciate the diversity of engineering environments and the different skill sets and mindsets they require.

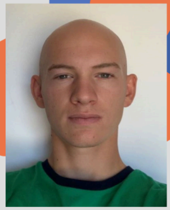
9.2 Personal and Professional Growth Through Internship Experience

This internship was a highly valuable experience, both technically and personally. I found the project genuinely exciting, especially the opportunity to work with cutting-edge technology like DeepMind’s GenCast. Even though the final results didn’t fully meet expectations—due to time constraints and the model’s current limitations—I still felt like I was contributing to a field that is rapidly evolving. It was rewarding to work on something that felt new and exploratory, where not everything had been done before.

One of my goals was to experience what it’s like to work in a large company. This internship gave me that exposure. I also gained confidence in my ability to work in an English-speaking professional setting, which reassured me that I can adapt and contribute in international environments.

Now that my studies are complete, I’m actively looking for a full-time position. This internship helped clarify my interests: I want to continue working in the field of AI, particularly in areas related to agentic AI—a direction I became increasingly interested in during the project. I’ve realized that I’m especially motivated by the opportunity to see a product or project grow over time, to witness its evolution and contribute to its continuous improvement. For my next step, I’m hoping to join a smaller company or startup, where I can benefit from closer mentorship, more collaborative technical growth, and a stronger sense of ownership over the work I contribute to.

Overall, this internship was a pivotal step in my transition from student to professional, and it has given me both the confidence and clarity to pursue my next career move with purpose.



JACQUES MEYER

ENGINEER STUDENT

CONTACT ME!

jacques.meyer@edu.ece.fr
+33 7 88 07 07 78
LinkedIn : Jacques Meyer

French Nationality
British Nationality
Driver's License Class B

PROFESSIONAL SKILLS

- C, C++, C#
- Java
- JavaScript, Node.js
- Python (OpenCV, Scikit-learn, PyTorch, TensorFlow)
- R
- Spark, Hadoop ecosystem
- SQL , NoSQL
- Kubernetes, Docker
- CI/CD Pipeline
- Arduino, VHDL
- Linux

Other IT tools I have worked with :

- GitHub, Git
- JetBrains Toolwork
- Power BI
- Visual Studio Code
- AWS (AI)

LANGUAGES:

- French (Native)
- English (C1) - 930 TOIEC
- German (B2)

SOFT SKILLS

- Committed to deliver
- Motivated
- Team spirit
- Organized
- Curious

ABOUT ME

As part of my studies, I am looking for a 6 **month internship** in the **domain of Data Science and Artificial Intelligence**, from **March 2025**.

I am eager to learn, passionate about AI and machine learning, with keen analytical and problem-solving abilities.

EDUCATION

Engineering Diploma, Master of Science in AI and Big Data(ECE Paris)
2020 - 2025 (present)
- 5th year MSc Engineering student. Specializing in AI & Information Systems
- Coursework includes Machine/Deep Learning, Data Analytics , Advanced Databases, Cloud, Big Data, Information Security & Networking
- Rank 9/100, 4/100, 11/100 in the last 3 semesters

- **6 months semester abroad in Thailand** (Royal University Chulalongkorn Bangkok)
  - primary focus on Digital Communication & User Interface and Experience (UI/UX)

French Scientific Baccalauréat & German Abitur
Bilingual Lycée Jean Mermoz, St-Louis, FR ; ABIBAC section (French/German)
2016 - 2019

RELEVANT EXPERIENCE

End-of-Study Project *(ECE Paris)*
2024 – Present
- *Conducting a data-driven research project focused on analyzing the impact of stress and cognitive fatigue on students' academic performance. Leveraging Deep Learning techniques with audio reccordings and advanced data analysis to uncover patterns and insights.*

Internship - French Army's Technical Department *(STAT)*
2024 April to August ; (Python, OpenCV, Tkinter, Yolo)
- *Enhancement of an anti-drone system on a 20mm cannon by integration of a thermal camera and weather sensors, focussing on software development for the user interface, AI detection on videos, and ballistics calculations.*

Project at the French National Institute of Sport *(INSEP)*
2023 - 2024 ; (Python, Dart)
- *Development of an application for collecting and analyzing elite badminton athletes' body movements using accelerometer sensors, aimed at enhancing game efficiency.*

Student Projects *(ECE Paris, FR)*
2020 - present
- *"Hands on" project-based learning examples:*
  - MachineLearning, DataAnalytics and DeepLearning projects (python and R) , *Creation of a a fully autonomous plane simulator based on graph theory in java , a Scrabble game in C , a Cluedo game in C++, a instant messaging app using Java.*

Student Tutor (Tutoring association of ECE Lyon, FR)
2021 - 2022
- Programing, Physics, Maths

JACQUES MEYER

ENGINEER POSTGRADUATE

CONTACT ME!

jacques.meyer@edu.ece.fr
+33 7 88 07 07 78
LinkedIn : Jacques Meyer

French Nationality
British Nationality
Driver's License Class B

PROFESSIONAL SKILLS

- C, C++, C#
- Java
- JavaScript, Node.js
- Python (OpenCV, Scikit-learn, PyTorch, TensorFlow)
- R
- Spark, Hadoop ecosystem
- SQL , NoSQL
- Kubernetes, Docker
- CI/CD Pipeline
- Arduino, VHDL
- Linux

Other IT tools I have worked with :

- GitHub, Git
- JetBrains Toolwork
- Power BI
- Visual Studio Code
- AWS (AI)
- Jira

LANGUAGES:

- French (Native)
- English (C1) - 930 TOIEC
- German (B2)

SOFT SKILLS

- Committed to deliver
- Motivated
- Team spirit
- Organized
- Curious

ABOUT ME

Eager to begin my professional career, I am looking for a **full-time role** in Data Science and Artificial Intelligence starting **January 2026**

I am eager to learn, passionate about AI and machine learning, with keen analytical and problem-solving abilities.

EDUCATION

Engineering Diploma, Master of Science in AI and Big Data (ECE Paris)
2020 - 2025 (present)
- 5th year MSc Engineering student. Specializing in AI & Information Systems
- Coursework includes Machine/Deep Learning, Data Analytics , Advanced Databases, Cloud, Big Data, Information Security & Networking
- Rank 9/100, 4/100, 11/100 in the last 3 semesters

- **6 months semester abroad in Thailand** (Royal University Chulalongkorn Bangkok)
  - primary focus on Digital Communication & User Interface and Experience (UI/UX)

French Scientific Baccalauréat & German Abitur
Bilingual Lycée Jean Mermoz, St-Louis, FR ; ABIBAC section (French/German)
2016 - 2019

RELEVANT EXPERIENCE

End of Study Internship *(Cognizant London)*
2025 April to September ;
- *Developed a weather forecasting system using DeepMind's GenCast model, focusing on deployment and optimization in AWS. Built GenAI agents to generate customizable, TV-style video forecasts for user-defined locations such as wind farms.*

End-of-Study Project *(ECE Paris)*
2024 – 2025 ;
- *Conducting a data-driven research project focused on analyzing the impact of stress and cognitive fatigue on students' academic performance. Leveraging Deep Learning techniques with audio recordings and advanced data analysis to uncover patterns and insights.*

Internship - French Army's Technical Department *(STAT)*
2024 April to August ;
- *Enhancement of an anti-drone system on a 20mm cannon by integration of a thermal camera and weather sensors, focusing on software development for the user interface, Computer Vision (OpenCV, Yolo), and ballistics calculations.*

Project at the French National Institute of Sport *(INSEP)*
2023 - 2024 ;
- *Development of an application for collecting and analyzing elite badminton athletes' body movements using accelerometer sensors, aimed at enhancing game efficiency.*

References:

- Copernicus API :
  https://cds.climate.copernicus.eu/datasets/reanalysis-era5-pressure-levels?tab=download

- Deepmind forum :
  forecasting in real time with GenCast_mini_demo · Issue #131 · google-deepmind/graphcast

- Deepmind Era5 Bucket (recommended by Deepmind searcher) :
  google-research/arco-era5: Recipes for reproducing Analysis-Ready & Cloud Optimized (ARCO) ERA5 datasets.

  WeatherBench 2 Data Guide — WeatherBench 2 documentation

- Deepmind's public GenCast model :
  https://github.com/google-deepmind/graphcast?tab=readme-ov-file

- Deepmind's GenCast research paper :
  https://www.science.org/doi/10.1126/science.adi2336

- Wind Turbine information :
  https://www.thewindpower.net/turbine_en_20_siemens_swt-3.6-107.php

Evaluation Form « COMPANY »
To be completed by training tutor
To insert in the internship report

Internship Evaluation Form
Ing5 – 3rd year Engineering Cycle
2024/2025

Student: **Jacques Meyer** Major: **DATA IA**
Company: **Cognizant**
Training tutor: **Neel Savani**
Tutor's ☎: **+44 7490 999017** Tutor's email : **Neel.Savani@cognizant.com**

Possibility of granting ½ point	**Note**
Mastery of science and technology • Analytical skills/ understanding of issues • Implementation of knowledge • Ability to acquire new knowledge	10/10
Mastery of methods and tool of the engineer • Methodology/ work organization, project management : development of a tool or methodology • Synthesis and communication of results, mastery of communication tools	10/10
Conduct of action and decision making • Achievement of objectives, quality of work, compliance with specifications • Autonomy/initiative/creativity/opening/motivation • Compliance with procedures	10/10
Integration into an organization and ability to conduct • Ability of integration : to express expectations, to speak in a debate, ability to listen, to accept criticisms and to question themselves • Communicate about activities and make reports • Consideration of economic and business issues	5/5
Respect for societal values, social and environmental • Appropriation of values, codes, team culture and organization • Ethical behavior	5/5
TOTAL	**40 /40**
TOTAL	20 /20

Observations :

Jacques has demonstrated exceptional performance during his internship, exceeding expectations across all evaluated areas. His project focused on leveraging AI, specifically Generative AI, for weather prediction within the logistics industry, a complex and impactful undertaking.

Jacques exhibited remarkable technical proficiency, seamlessly applying his understanding of AI principles to practical applications. His analytical skills were outstanding; he adeptly identified critical issues and approached challenges with insightful critical thinking. Jacques's ability to translate theoretical knowledge into tangible deliverables was evident in the successful outcomes of his weather prediction models.

Notably, Jacques showcased a proactive and rapid capacity for acquiring new knowledge, quickly mastering technical

leadership by effectively guiding and motivating a team of fellow interns. Jacques's organizational skills and project management approach were exemplary.

His communication of complex AI concepts and results was clear, concise, and highly effective. Jacques consistently met objectives with work that exceeded expectations, adhering strictly to my specifications. He displayed a high degree of autonomy, initiative, and creativity, proactively seeking solutions and contributing innovative ideas. Jacques integrated seamlessly into our team, embodying our values and demonstrating excellent ethical conduct.
Jacques has earned full marks for his outstanding internship performance.

At Cognizant London, the 29 July 2025
Signature of the tutor
(compulsory)

cognizant
Stamp of the host company
(copulsory)