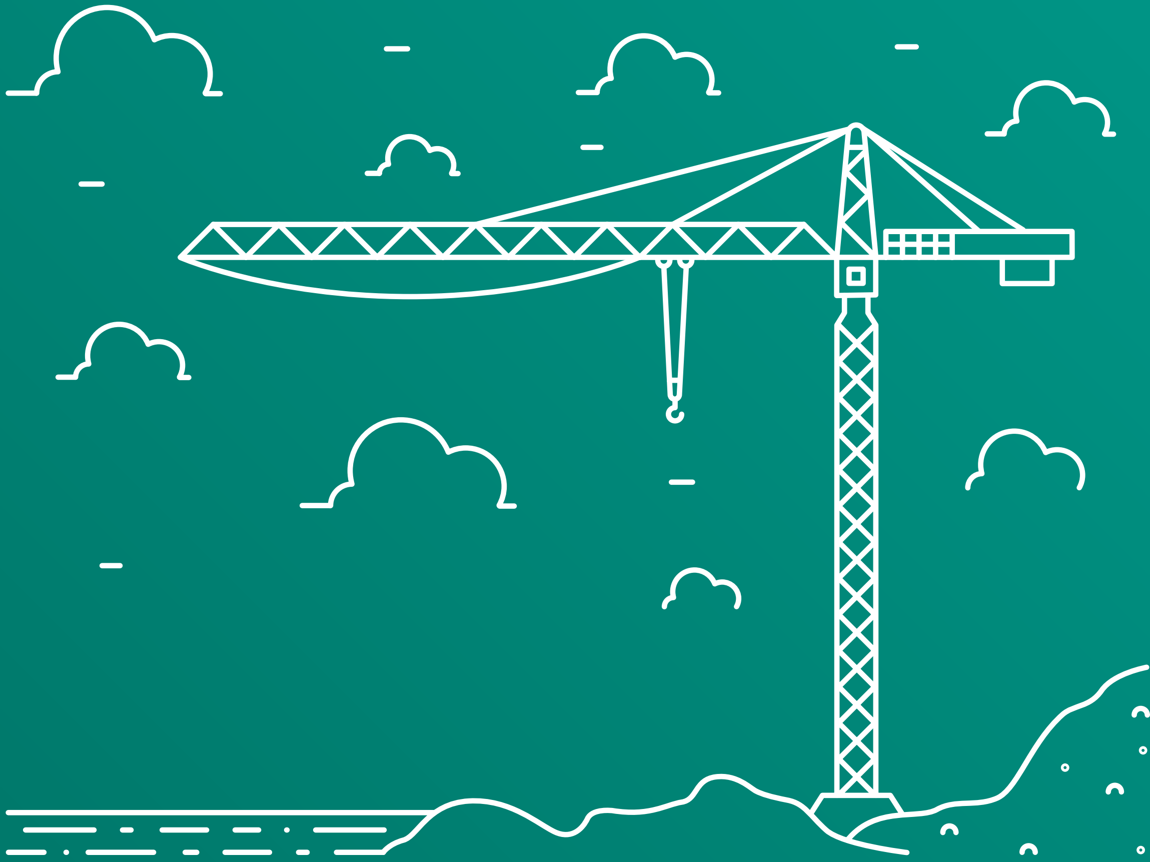
Hoist

WHITEPAPER

# Innovating safely with AI



> “AI has the potential to improve nearly every aspect of people’s lives but it also creates serious risks.”

**SAM ALTMAN, CEO OF OPENAI**

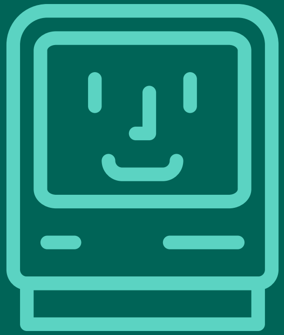
## Contents

AI HAS EXPLODED ON TO THE BUSINESS SCENE, AND WHILE IT OFFERS GREAT POTENTIAL, WE NEED TO TREAD CAREFULLY AS WITH ALL NEW TECHNOLOGY.

Our staff should all be more productive by augmenting their work with AI, but how do we roll this out across a company without compromising our confidential data and managing the cost along the way?

In this whitepaper, we look at some aspects of AI that corporates should be aware of, we review an AI policy, and we look at an AI sandbox that would meet that policy while opening the door to innovation within your business.

For the purpose of this whitepaper "generative AI" is any program that is capable of generating content that uses a training set as a method of defining the patterns it will generate.



## IT’S NOT MAGIC

Generative AI can seem magical or human, but really it is like a calculator: it can do complex math very quickly but does not understand maths at all, and is just following rules given inputs.

Generative AI differs from a calculator in that the rules it follows are derived from huge datasets using statistical models, making the number of these rules hard for a human to comprehend, and seemingly unknowable.

## IT IS A CLOUDY MIRROR

The thing to know about this generation method is that it is a cloudy mirror of the data it was trained on. This means that the generation has the same bias as the training set.

A training set trained with only positive things to say about a subject is unlikely to say negative things about that same subject, and therefore will not be able to give a balanced argument on the subject.

## TRUST BUT VERIFY

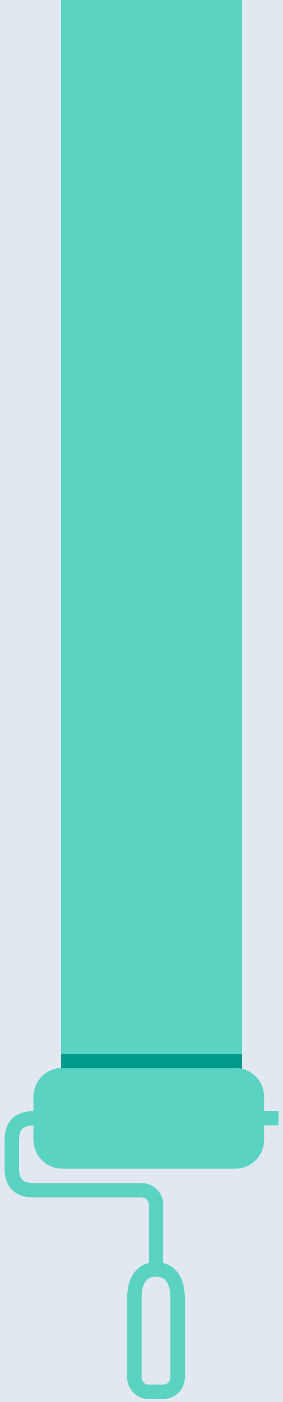
Most models have the ability to “hallucinate” or create content from existing patterns in the training set, not just babble back the training set. Which means that the content that they give you may not be true, for example it may put a sentence where it shouldn’t be. In fact it may not be real, as it can generate content of the same form as the truth, but which never existed. The classic example of this is generating believable references to texts that do not exist.

## IT DOESN’T GENERATE ANYTHING “NEW”

If something new is generated it is because of you and not the AI.

While the AI can generate new sequences of words (or pixels on a screen) in the same way that rolling a dice generates a new sequence of numbers, it does not have the ability to create new knowledge.

It is through direction and editing that you will be able to focus it to create something new.



## BEWARE THE TRAINING DATA

The training set may be unknown and AI could be creating a copy of someone else’s work or a very similar version. This may also be true if it has the ability to search and use resources as part of its generation.

As it may not tell you what it used there is a possible copyright infringement without your knowledge, or a hallucination of those references.

## STORING AND USAGE OF DATA

As the goal of most AI is to make it more effective, it makes sense to retrain the AI on user data, however if we were to use our IP then it would be incorporated into this training set for use by other people.

A secondary concern is if the data sent to the AI is kept in accordance with security and privacy concerns.

## INACCURACIES AND BIAS

Maybe a good way to think of generative AI is to imagine a parrot that has heard lots of words as answers to questions and can repeat them, but has no understanding of those answers. So when you ask it a question it will do its best to parrot back an answer that it thinks sounds like the ones it has heard before.

Given this example it is obvious that you would check the parrots work and not take it at face value, and maybe ask who it has been listening to. Was it in the company of too many salty pirates? If so it's answers maybe skewed towards the ideas and ethos of those pirates.

Content generated by an AI must be checked for accuracy and unintended bias.

# AI risks

| RISK | DO | DON’T |
|---|---|---|
| Privacy | Understand the T&Cs of the services you are approving to know what data has been used for training and whether the prompts you provide are being used for training.<br>Stand-up your own AI via API in order to use confidential information for training and/or prompts. | Enter confidential information into a personal or public AI, even within a paid subscription.<br>Use confidential information to train an AI that is going to be used by people who should not have access to the confidential information. |
| Access | Ensure your team are using single-sign-on (SSO) to access AI. | Use personal accounts. |
| Trust | Review the work produced from an AI in the same way you would review the work from a human. | Assume the information you get is correct just because it’s well written. |
| Innovate | Encourage your staff to explore AI as part of their normal work using approved AI services. | Assume it will replace employees.<br>Avoid using it because its new and because you can’t see the benefit yet. |
| Manage costs | Keep track of the extra costs associated with AI and ensure you aren’t paying multiple times for the same service just because it’s available everywhere. | Assume that every tool you use that offers AI is going to add value. |

# AI policy

AN AI POLICY SHOULD COVER ANY USE OF AI TOOLS AT WORK FOR ALL STAFF, UNLESS THEY HAVE BEEN GIVEN A SPECIFIC EXCLUSION.

A basic policy can start with an overview of things that your team should be aware of when using AI and then be followed by a specific list of approved tools and restrictions.

## APPROVED TOOLS

Provide a list of approved AI tools overseen by the CIO, having three classifications:

### GENERAL BUSINESS

Approved for business in confidence and public data

### CONFIDENTIAL

Approved for use for IP (including designs, code, etc) and confidential information

### RESTRICTED

Approved for use of personal data (both common and sensitive)

## POLICY RESTRICTIONS

You MUST only use approved AI tools.

You MUST understand the data that you sending to the AI and how that fits to the classification above.

You MUST check the result given by the AI for accuracy and unintended bias.

You SHOULD mark substantive generated work as such if it hasn't been edited by a human.

# Hoist: a white-label platform

A WHITE-LABEL PLATFORM* PROVIDES A RING-FENCED, CONTROLLED ENVIRONMENT FOR YOUR COMPANY TO CONSUME AI TOOLS AND START YOUR OWN JOURNEY OF INNOVATION.

*I.e. https://ai.yourdomain.com

1.

# START WITH A CONTROLLED ENVIRONMENT

A white-label platform backed by robust policies and with access via single-sign-on is a security-first solution.

It allows you to provide clear policy guidance to your staff, control the AI services made available and get reporting on usage while protecting privacy and confidential information.

## 2.

## ENCOURAGE YOUR TEAM TO START EXPLORING WITH CLEAR BOUNDARIES

To explore the question of what is possible, we need a safe way to experiment. A platform enables your staff to start experimenting with AI tools in their daily jobs, using confidential information, trying different prompts and using tools to verify and visualise their information.

Initially, you can provide your staff with access to ChatGPT and your own trainable generative AI.

These will be provided using your API keys, so only you have access and you will pay the direct usage costs with no markup.

## 3.

## R&D

Experimentation is likely to lead you to wanting some R&D, which can be prohibitively expensive.

Having a platform, which your staff already have access to, which is already connected to AI services is an ideal starting point.

From here, the EndGame team can help you develop your own IP on the platform.

Hoist

# Case study

ENDGAME IS A SOFTWARE AGENCY IN NEW ZEALAND AND ALONGSIDE OTHER PROFESSIONAL SERVICE COMPANIES, WE ARE ASKING OURSELVES:

**Are the recent advancements in AI going to be like Kodak's downfall?**

**Will our "time and materials" business model be disrupted as service delivery time decreases?**

Recently one of our engineers commented that they're glad to live in a thin slice of history where they can write code for a living. While we don't see this disappearing any time soon, the real question we need to ask ourselves is:

**What should we do about this now?**

# AT ENDGAME, WE ARE RESPONDING IN THREE WAYS:

## 1.

### Providing “foundation” access to AI tools

So every employee can experiment within a controlled environment, sure that confidential or customer information isn’t leaked.

## 2.

### Building knowledge bases

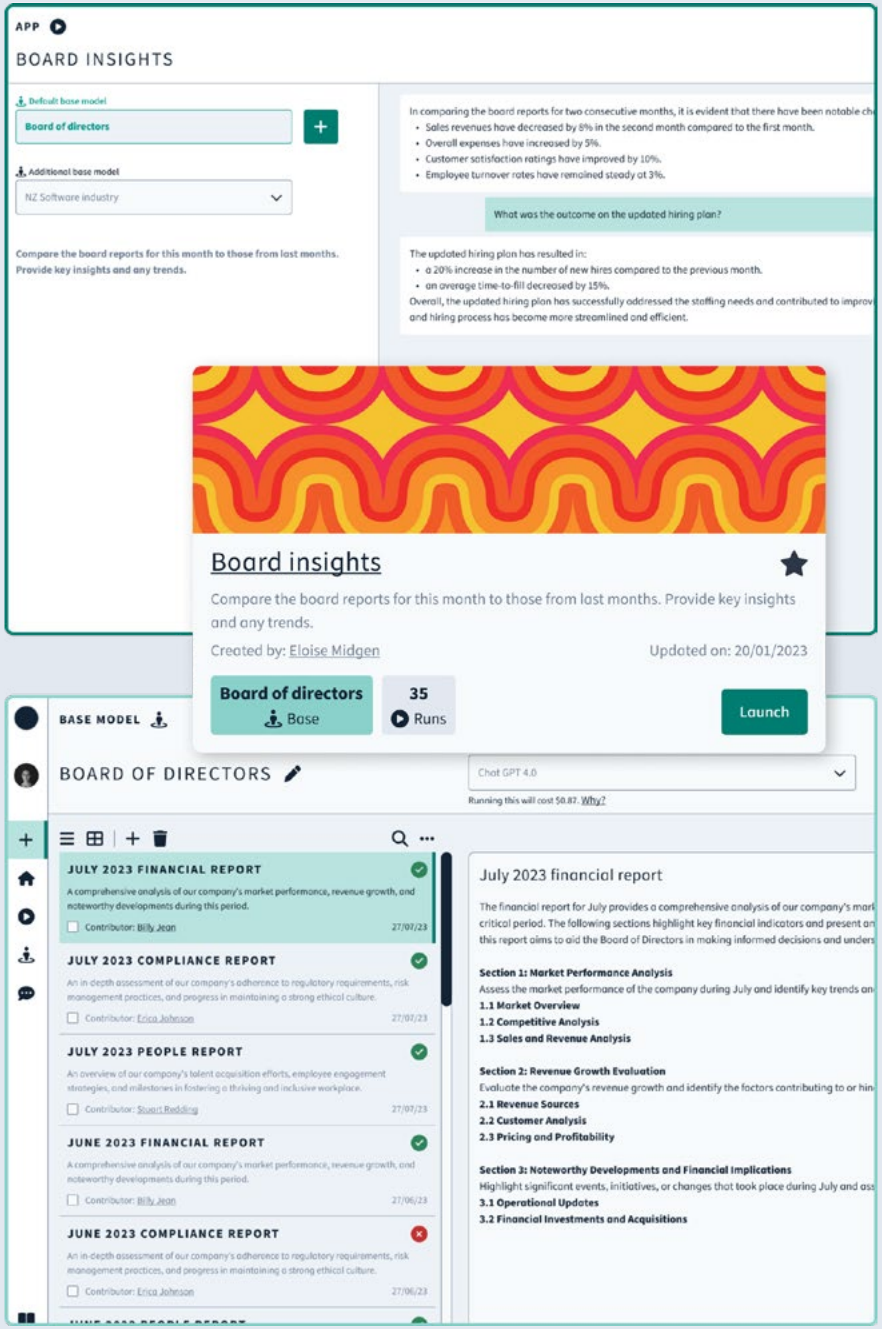
These “base models” are built with the accumulated context and know-how for each client, project and function within the company and allow us to use AI in a way that is more relevant and useful for each client. Our goal is to offer more value to our clients using AI than they can get on their own using AI.

## 3.

### Helping our team learn how to use AI.

So teams can write prompts that use AI as more than just a faster search engine. We are encouraging our team to write and share “apps” with each other that use our company and client base models.

We created **Hoist** to help us and others like us on this journey. To the right is an example of Hoist using our Board Reporting base model. Our Board of Directors can accumulate board papers and learn how to use AI apps to help with analysis and trends. All in a controlled environment which ensures we protect our confidential information.



BOARD INSIGHTS

Default base model: Board of directors

Additional base model: NZ Software industry

Compare the board reports for this month to those from last months. Provide key insights and any trends.

In comparing the board reports for two consecutive months, it is evident that there have been notable ch
- Sales revenues have decreased by 8% in the second month compared to the first month.
- Overall expenses have increased by 5%.
- Customer satisfaction ratings have improved by 10%.
- Employee turnover rates have remained steady at 3%.

What was the outcome on the updated hiring plan?

The updated hiring plan has resulted in:
- a 20% increase in the number of new hires compared to the previous month.
- an average time-to-fill decreased by 15%.

Overall, the updated hiring plan has successfully addressed the staffing needs and contributed to improv and hiring process has become more streamlined and efficient.

Board insights

Compare the board reports for this month to those from last months. Provide key insights and any trends.

Created by: Eloise Midgen

Updated on: 20/01/2023

Board of directors Base | 35 Runs

BASE MODEL

BOARD OF DIRECTORS

Chat GPT 4.0

Running this will cost $0.87. Why?

JULY 2023 FINANCIAL REPORT ✓
A comprehensive analysis of our company’s market performance, revenue growth, and noteworthy developments during this period.
□ Contributor: Billy Jean 27/07/23

JULY 2023 COMPLIANCE REPORT ✓
An in-depth assessment of our company’s adherence to regulatory requirements, risk management practices, and progress in maintaining a strong ethical culture.
□ Contributor: Erica Johnson 27/07/23

JULY 2023 PEOPLE REPORT ✓
An overview of our company’s talent acquisition efforts, employee engagement strategies, and milestones in fostering a thriving and inclusive workplace.
□ Contributor: Stuart Redding 27/07/23

JUNE 2023 FINANCIAL REPORT ✓
A comprehensive analysis of our company’s market performance, revenue growth, and noteworthy developments during this period.
□ Contributor: Billy Jean 27/06/23

JUNE 2023 COMPLIANCE REPORT ✗
An in-depth assessment of our company’s adherence to regulatory requirements, risk management practices, and progress in maintaining a strong ethical culture.
□ Contributor: Erica Johnson 27/06/23

July 2023 financial report

The financial report for July provides a comprehensive analysis of our company’s mark critical period. The following sections highlight key financial indicators and present an this report aims to aid the Board of Directors in making informed decisions and unders

**Section 1: Market Performance Analysis**
Assess the market performance of the company during July and identify key trends an
**1.1 Market Overview**
**1.2 Competitive Analysis**
**1.3 Sales and Revenue Analysis**

**Section 2: Revenue Growth Evaluation**
Evaluate the company’s revenue growth and identify the factors contributing to or hin
**2.1 Revenue Sources**
**2.2 Customer Analysis**
**2.3 Pricing and Profitability**

**Section 3: Noteworthy Developments and Financial Implications**
Highlight significant events, initiatives, or changes that took place during July and ass
**3.1 Operational Updates**
**3.2 Financial Investments and Acquisitions**

# Hoist

Powered by EndGame

Trusted cloud software providers of

ACC He Kaupare. He Manaaki. He Whakaora. prevention. care. recovery.

bnz

MASSEY UNIVERSITY
TE KUNENGA KI PUREHUROA
UNIVERSITY OF NEW ZEALAND

Find out more: hoist.io