TEAM GENEX - IDEA SUBMISSION

Business Focus Area: Improve Safety under Energy and Resources area

Team Members: (TISS Mumbai, Masters in Analytics program)

Rahul Nain

Problem statement:

"Reporting accidents in mining and energy sector inv
leads to data procrastination for procedures such as
incidence, approval of medical insurance and compliance reporting"

Business Challenge and Opportunity

Global Market Size Workplace Safety in mining & related industries
USD 14.2 billion

Expected Growth
USD 26.7 billion By 2027

Indian market size for workplace safety & risk mitigation in mining and related industries
USD 2.8 billion

Expected Growth
USD 5.4 billion By 2028

Business Challenge - Competitor Analysis

Competitor	Key Features and Focus	Strengths	Weaknesses
HexaSafe Solutions	Real-time hazard monitoring and reporting	Established player, real-time monitoring	Limited AI integration for predictive risk mitigation
SafeTech Insights	Advanced safety analytics, data-driven insights	Strong analytics foundation	Less intuitive UI compared to potential AI-driven solution
PredictRisk Innovations	Predictive modeling of workplace hazards	Expertise in predictive modeling	Limited automation for hazard reporting and response
RiskMinder AI	AI-driven insights, customizable hazard identification	AI-driven approach for dynamic mitigation	Potential learning curve for users unfamiliar with AI

Business Opportunity - Saving cost and time

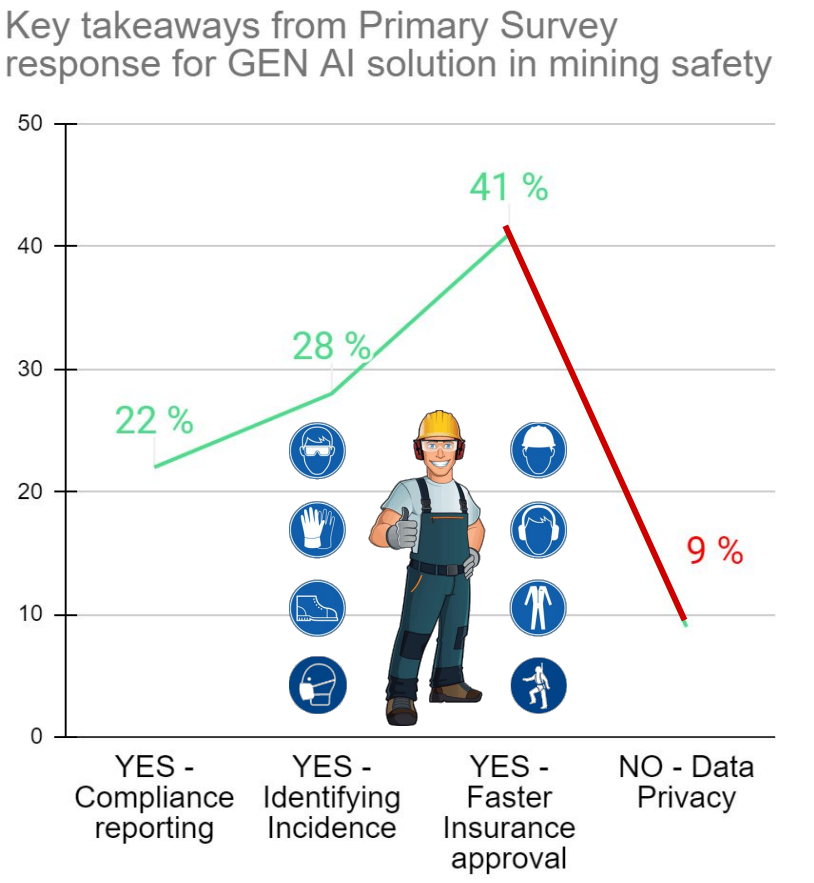
- Annual Loss of Lives: Approx. 2 million lives lost annually.
- Economic Cost: Estimated at $1.25 trillion USD.
- Cost of Occupational Injuries: One study reports a staggering $3 trillion.
- Inefficient Safety Reporting Systems: Businesses in the U.S. lose an average of $1.8 billion yearly.
- Global Economic Impact: Work-related injuries cost ranges from 1.8% to 6% of GDP.
- Lost Workdays: Total of 31,515,368 lost workdays over the period.
- Equivalent to 5,700 person-years lost per year

Source: ILO NCBI NSC NCBI

.

Rationale behind GEN AI Solution and Technologies Proposed

Primary Research Survey



There is a need for GEN AI solution keeping in check Data Privacy Risk

Secondary Research Data

Aurum Mining Corporation Data set has been used to train our prototype using:

- Input: Incident Narrative
- Output: Classification into column variables

Business Use cases:

- Incident Analysis to identify root cause
- Fast tracking medical care and health insurance
- Compliance Reporting

Available Technologies

Chosen because it is really good at understanding and generating human-like textl

GPT (Text-to-Text Transfer Transformer)

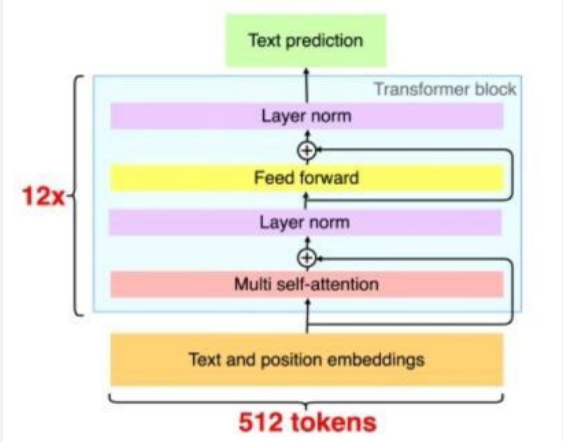
GPT, which stands for "Text-to-Text Transfer Transformer," is a highly versatile language model renowned for its adaptability in transforming various text inputs into meaningful text outputs. This adaptability makes it well-suited for a wide range of language generation tasks.

.

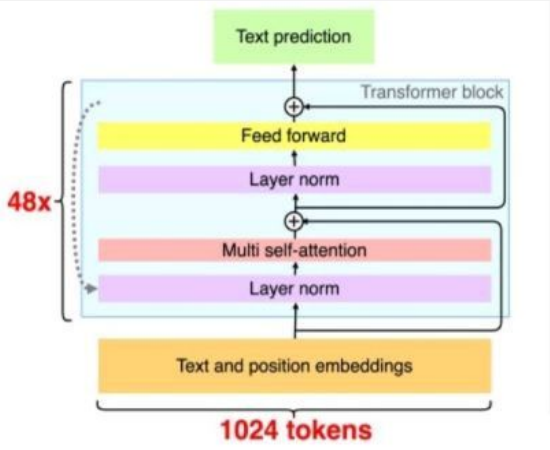
Key Features of T5

- Text-to-Text Framework
- Fine-Tuning Flexibility
- Competitive Performance
- Interpretable Outputs
- Transfer Learning

GPT1



GPT2



Proposed Generative AI solution including Solution Architecture

Architecture of Transfer Learning

The data set contains narrations of incidents and its **manual** classification of different attributes of the incident

Fine-Tuning

T5 Base Transformer

T5

Python Preprocessing to obtain data in excel format

Faster preprocessing of data and reporting of incidents

Reference Research Paper Link

Example of Narration :

Employee was building and plastering a brattice when some of the plaster got on he's glasses and into he's left eye causing scratches and irritation to he's eye.

INPUT TEXT

Classification Output :

Column 1
Subunit : Underground
Underground location : Vertical Shaft
Underground mining method : Continuous Mining
…..

Column 2
Body Part Injured : EYE/OPTIC NERVE
…..

EXPECTED OUTPUT

Applications:

- Incident Analysis to identify root cause
- Fast tracking medical care and health insurance
- Compliance Reporting

Proposed Business Benefits

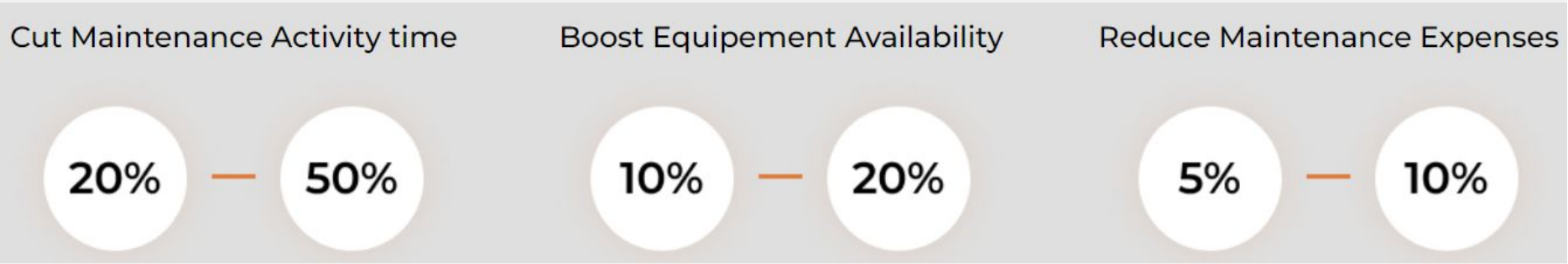
Revenue Increase: Streamlined incident analysis accelerates identification of patterns

Can save estimated losses due to direct and indirect cost of around **1.2 Bn $ Annually** per person.

Cost Reduction: Swift root cause identification minimizes recurring incidents and costs.

Workplace safety has great ROI for **$4 to $6 for every $1 invested in workplace safety (OSHA).**

Improve Asset Utilization: Accurate incident analysis enables proactive maintenance, lowering downtime, increase equipment availability and reduced maintenance costs.



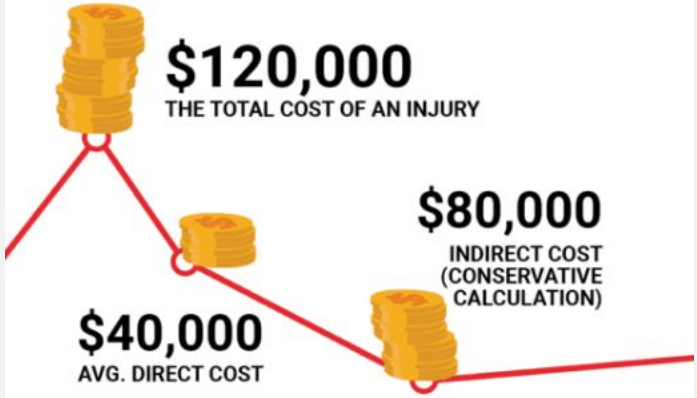
Risk Reduction: Enhanced compliance minimizes legal and financial risks and potential penalties

The maximum **penalties** will be raised from the current **$100,000** and/or up to one year imprisonment

Experience Enhancement: Effective incident response fosters a safer, positive work environment.

Organizations prioritizing safety lower safety incidents by **70%** according to a study by **Gallup 2017.**

The **NSC** estimated a total of **99,000,000 days lost** due to work-related injuries for 2020.



$4 - $6 AVG. ROI

FOR EVERY DOLLAR INVESTED INTO WORKPLACE SAFETY

50,000,000

DAYS LOST IN FUTURE YEARS FROM 2020 IN JURIES

Risks associated with proposed solution -> Mitigation Strategies

- Data Privacy -> Data Encryption
- Proliferating Devices -> Data Governance
- Maintenance Costs -> Automation
- Access Controls -> Better Data Architecture
- Visibility of all Data -> Effective Data Pipeline implementation
- Bad Data Culture or Data Bias -> Data Ethics trainings and policies
- Scalability of solution - > Developing Quality standards
- Regulations -> Staff Training

PRODUCT DEVELOPMENT PRESENTATION

Business Focus Area: Improve Safety under Energy and Resources sector

TEAM GENX (TISS Mumbai, Masters in Analytics program)

Soumyadeep Pal

Amit Kadarmandalgi

Mayuri Jape

Rahul Nain

TECHNOTHRIVE 23
EMPOWERING CREATORS

YASH Technologies
More than what you think.

unstop

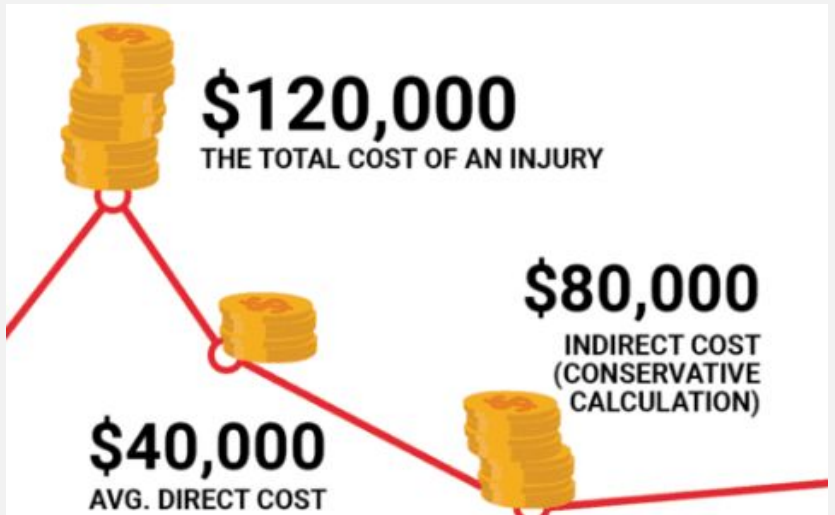
RE-IMAGINING FUTURES
TISS

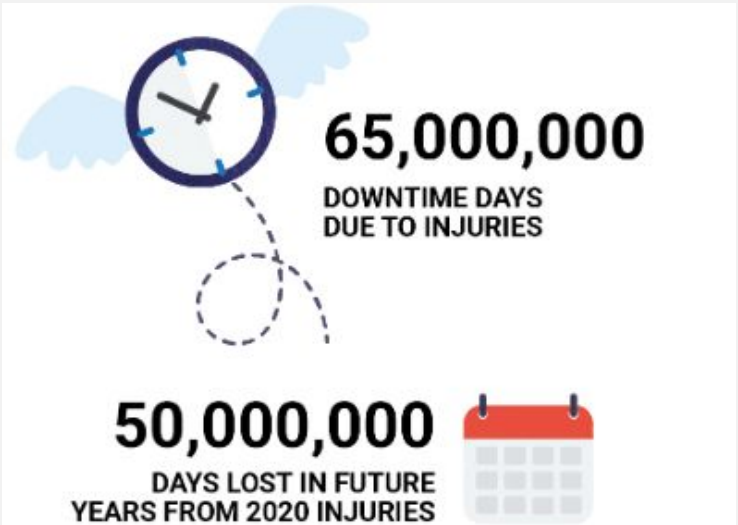
Business Problem Statement & Scope of Solution

“Reporting accidents in mining and energy sector involves manual data entry which leads to data procrastination for procedures such as identifying root cause of incidence, approval of medical insurance and compliance reporting”

- **What is the business problem?** Manual entry of incident reports of accidents happening at workplace
- **Where is this happening?** Mainly in energy and mining sector where there are higher chances of work hazards
- **How did we arrive at this problem?** Conducted a primary research using a questionnaire based survey form
- **Why is this an important issue?** Manual entry leads to data procrastination which leads to delay in processes.
  - **Root Cause analysis:** Not identifying the cause of accidents implies no corrective or preventive action
  - **Insurance approval:** Delay in approvals for crucial medical assistance and insurance approval to the victim
  - **Regulatory Compliance:** Failing to do internal audits can lead to legal penalties also imprisonment.

Compromising on safety standards of workers leads to loss of **life, money, time** and also **legal punishments**





Workflow Solution and Technical Architecture

Architecture of Transfer Learning

The data set contains narrations of incidents and its **manual** classification of different attributes of the incident

Fine-Tuning

GPT2 Transformer

Reference Research Paper Link

Python Preprocessing to obtain data in excel format

Faster preprocessing of data and reporting of incidents

Applications:

- Incident Analysis to identify root cause
- Fast tracking medical care and health insurance
- Compliance Reporting

Example of Narration :

Employee was building and plastering a brattice when some of the plaster got on he's glasses and into he's left eye causing scratches and irritation to he's eye.

INPUT TEXT

Classification Output :

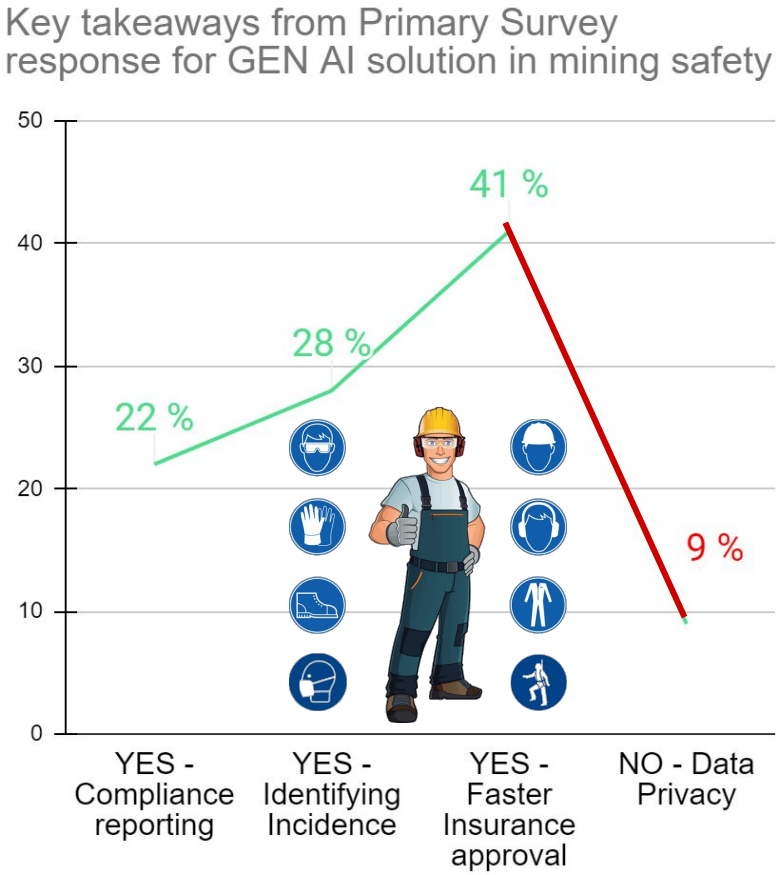
Column 1
Subunit : Underground
Underground location : Vertical Shaft
Underground mining method : Continuous Mining
…..

Column 2
Body Part Injured : EYE/OPTIC NERVE
…..

EXPECTED OUTPUT

Data Set, Sources and Data Processing Steps

Primary Research Survey Data



There is a need for GEN AI solution keeping in check Data Privacy Risk

Secondary Data Set Source

Data Set used:
Aurum Mining Corporation
Data set has been used to train our model using LLM

Data Type: Human data entries of Narratives of accidents in text form

Data Source: We have sent an official request to QHSE and received AMC safety data set after approval

[Safety Data Sheets - AMC Drilling Optimisation (amcmud.com)]

Data Processing Steps

1. Data Collection

Conducted Primary and Secondary research and Finalised on Data Set i.e. AMC Safety Data Set

2. Data Preprocessing

Data Pre processing to identify errors in text narrative, data selection and tokenization

3. Data Modelling

Splitting into training and test data set for model

Feedback and Corrections to enhance Model

Output

Details of foundation models

Choice of GPT (text to text transformer):

GPT2

Chosen because it is really good at understanding and generating human-like text

Key Features of GPT

1.Text-to-Text Framework: utilizes a text-to-text framework, enabling it to process input text and generate meaningful output text which supports a wide range of NLP tasks.

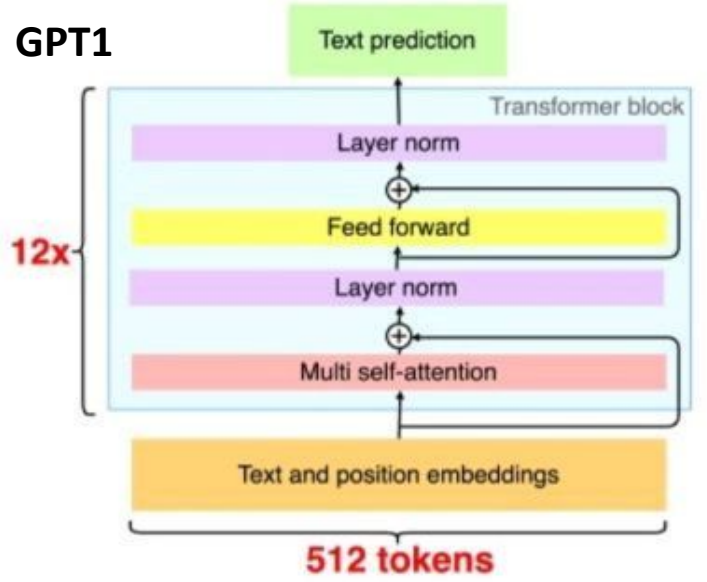
2. Fine-Tuning Flexibility: adaptable as it allows customization for specific tasks

3. Competitive Performance: consistently delivers competitive performance across various NLP benchmarks, thanks to its large-scale architecture and extensive pre-training.

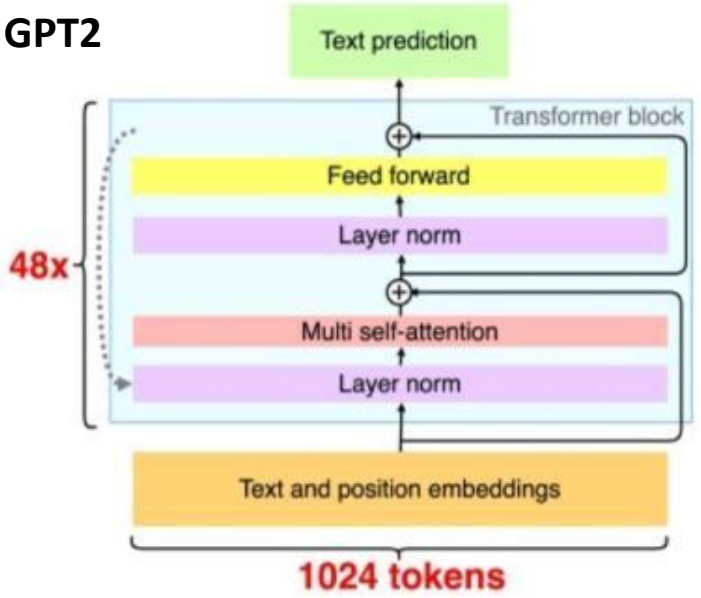
4. Interpretable Outputs: GPT generates interpretable and contextually meaningful outputs, aligning well with input context

5. Transfer Learning: leverages transfer learning by pre-training on vast text data, enabling it to understand language comprehensively.

GPT1



GPT2



Model evaluation & Deployment approach

Non-Traditional approach of model evaluation

1. Accuracy of Information: How accurately the generative AI model fills in the columns compared to manual entry.(Around 80% for the current model)
2. Time Efficiency: The time saved by using the generative AI model compared to manual data entry.(Need Company level data to calculate)

How should the metric look like?

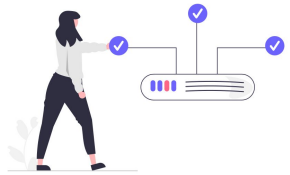
Example:
Manual Entry:
Accuracy: 90%
Time taken: 10 minutes per incident

Generative AI Entry:
Accuracy: 95%
Time taken: 2 minutes per incident

Model Deployment Approach

Link to GitHub Repository
Link to Demo Prototype

3 - Click Approach for front-end



Infrastructure Setup (Refer to Required Infrastructure and Environment)

1. Compute Resources (GPU recommended)

Adequate Storage
Stable Network Connection
Model Serving Setup

1. Choose Model Server Framework

Dockerize Application for Portability
API Development

1. Create RESTful API for Input and Response

Implement Robust Error Handling
Integration with Frontend/Client

5.Develop User Interface or Client App
Manage User Input and API Requests
Deployment and Monitoring

6. Host on Cloud Platform (e.g., AWS, GCP)
Implement Load Balancing (if needed)
Set up Monitoring, Logging, and Security Measures

GPT2 Fine-tuning Approach

Installing Dependencies
1.transformers
2.SentencePiece
3. Cython
5. rich

Choice of **platform** for fine-tuning.
Kaggle provides with 15GB of RAM and 15GB of GPU P100 Accelerator.

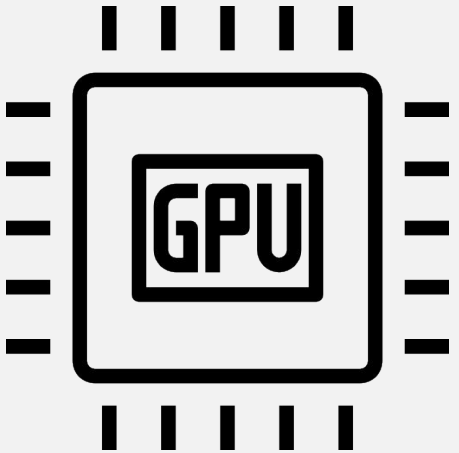
Choice of Model and libraries to Load the model: We have used GPT2 base model from Huggingface for finetuning.
Libraries: GPT2LMHeadModel and GPT2Tokenizer

```
from torch import cuda
device = 'cuda' if cuda.is_available() else 'cpu'

model = model.to(device)
```

```
def fine_tune_gpt2(model_name, train_file, output_dir):
    # Load GPT-2 model and tokenizer
    model = GPT2LMHeadModel.from_pretrained(model_name)
    tokenizer = GPT2Tokenizer.from_pretrained(model_name)
```

1. The initial function fine_tune_gpt2 takes three inputs: model_name, train_file, and output_dir.
2. It loads a pre-trained GPT-2 model and a compatible tokenizer using from_pretrained.
3. It checks if a GPU is available. If a GPU is available, it moves the model to the GPU for faster training.
4. It creates a training dataset using the TextDataset class, configured with the provided tokenizer and file path.
5. A data collator is created for language modeling tasks, with masking set to False.
6. Training arguments are set, including output directory, number of epochs, batch size, and save settings.
7. The model is trained using a Trainer object, which manages the training process.
8. The fine-tuned model is saved along with the tokenizer configurations to the specified directory.



Required Infrastructure and Environment

Estimated GPU Memory Requirement and Implementation Cost Analysis

Model Configuration:

- GPT Model: GPT-2 Base version (**851.2 MB**)
- Description Token Size: 30 tokens
- Batch Size: 8

GPU Memory Estimation:

- Formula: GPU Memory (GB) = Batch Size * Model Size (GB) * Tokenization Factor
- Calculation: GPU Memory (GB) = 8 * 0.8512 GB * 30 = **20.43 GB**

Cloud Service Costs for 1 Year:

- Azure: Approximately **$314.29 for 1 year**
- AWS: Approximately **$3,293.76 for 1 year**

AWS Cost Analysis

Detailed Estimate

Group hierarchy	Region	Service	Upfront	**First 12 months total**
My Estimate	India	Amazon EC2	3293.76	**$3293.76**

Azure Cost Analysis

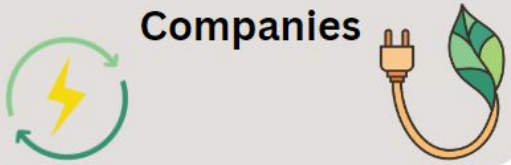
Service category	Region	Description	Estimated monthly cost
Compute	West India	1 D2 v3 (2 vCPUs, 8 GB RAM) (1 year savings plan), Linux, (Pay as you go); 0 managed disks – S4; Inter Region transfer type, 5 GB outbound data transfer from West India to East Asia	$77.77
Compute	East US	1 D2 v3 (2 vCPUs, 8 GB RAM) x 730 Hours (Pay as you go), Windows (License included), OS Only; 0 managed disks – S4; Inter Region transfer type, 5 GB outbound data transfer from East US to East Asia	$137.24
Compute	East US	1 A2 v2 (2 Cores, 4 GB RAM) x 730 Hours (Pay as you go), Windows (License included), OS Only; 0 managed disks – S4; Inter Region transfer type, 5 GB outbound data transfer from East US to East Asia	$99.28
Support	Support		$0.00
	Total		**$314.29**

Target Users

Mining Companies

Oil and Gas Companies

Renewable Energy Companies



Manufacturing and Construction Companies in Energy Sector

Regulatory and Compliance Bodies

Medical Personnel

Strategy Analysts & Decision Makers

Environmental Monitoring Officers

Risk Management and Insurance Agencies

Target User Personas

Operator Ojas

Operator Ojas is responsible for day-to-day operations in the mining and energy sector. They need tools that support their tasks and ensure safety.

Jobs To Be Done:

Job 1: Incident Notification and Response
In case of an incident, Oliver needs a tool that allows quick notification and response. The AI solution should facilitate immediate reporting and response procedures.

Compliance Officer Kartik

Compliance Officer Kartik ensures that the organization adheres to legal and regulatory requirements related to safety.

Jobs To Be Done:

Job 1: Automated Compliance Reporting
Carla's primary responsibility is to generate compliance reports. The AI solution should automate this process, providing accurate and up-to-date compliance reports.

Safety Manager: Sameer

Safety Manager Sameer oversees safety protocols and procedures within mining and energy operations. They are responsible for incident response and prevention strategies.

Jobs To Be Done:

Job 1: Efficient Incident Reporting
When an incident occurs, Sam needs to quickly report and respond to it. The AI solution should provide a seamless and streamlined reporting process, allowing Sam to receive real-time information and take immediate action.

Job 2: Root Cause Analysis
After an incident, Sam needs to identify the root cause to implement effective preventive measures. The AI solution should facilitate prompt identification and analysis of root causes.

Inspector Ishant

Inspector Ishant conducts routine inspections to ensure compliance with safety regulations and standards.

Jobs To Be Done:

Job 1: Real-time Compliance Reporting
Ingrid needs to ensure that the operations comply with safety regulations. The AI solution should provide instant access to compliance data, allowing Ingrid to make timely assessments and recommendations.

Medic Meera

Medic Meera handles medical assessments and treatments for injured personnel. Quick access to information is crucial for providing timely care.

Jobs To Be Done:

Job 1: Faster Approval of Medical Insurance
When an incident occurs, Mia needs to quickly approve medical insurance claims for injured personnel. The AI solution should expedite the approval process, ensuring prompt medical attention.

Analyst Alisha

Analyst Alisha analyzes safety data to identify trends, patterns, and areas for improvement. They rely on accurate and structured data for meaningful insights.

Jobs To Be Done:

Job 1: Clean and Structured Data
Alex needs high-quality, standardized data for effective analysis. The AI solution should provide clean and structured data, reducing the time and effort required for analysis.

Executive Evelyn

Executive Evelyn makes strategic decisions based on performance metrics and safety records. Informed decision-making is critical for operational efficiency.

Jobs To Be Done:

Job 1: Improved Decision-making
Evelyn relies on timely and accurate safety data to make informed decisions. The AI solution should provide access to such data for enhanced operational efficiency and risk management.

Business value proposition

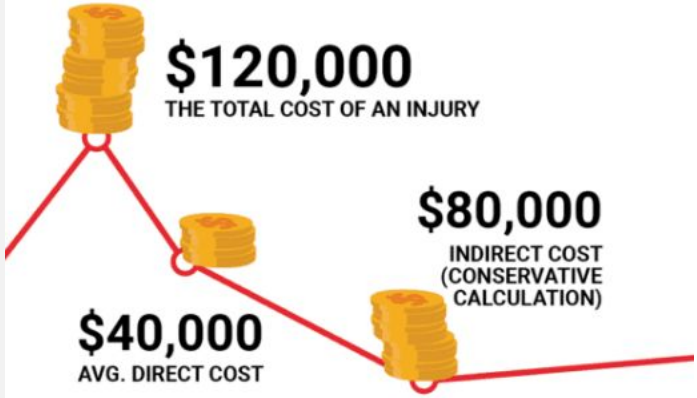
Revenue Increase: Streamlined incident analysis accelerates identification of patterns

Can save estimated losses due to direct and indirect cost of around **1.2 Bn $ Annually** per person.

Cost Reduction: Swift root cause identification minimizes recurring incidents and costs.

Workplace safety has great ROI for **$4 to $6 for every $1 invested in workplace safety (OSHA).**

Improved Asset Utilization: Accurate incident analysis enables proactive maintenance, lowering downtime, increase equipment availability and reduced maintenance costs.



Cut Maintenance Activity time	Boost Equipement Availability	Reduce Maintenance Expenses
20% — 50%	10% — 20%	5% — 10%

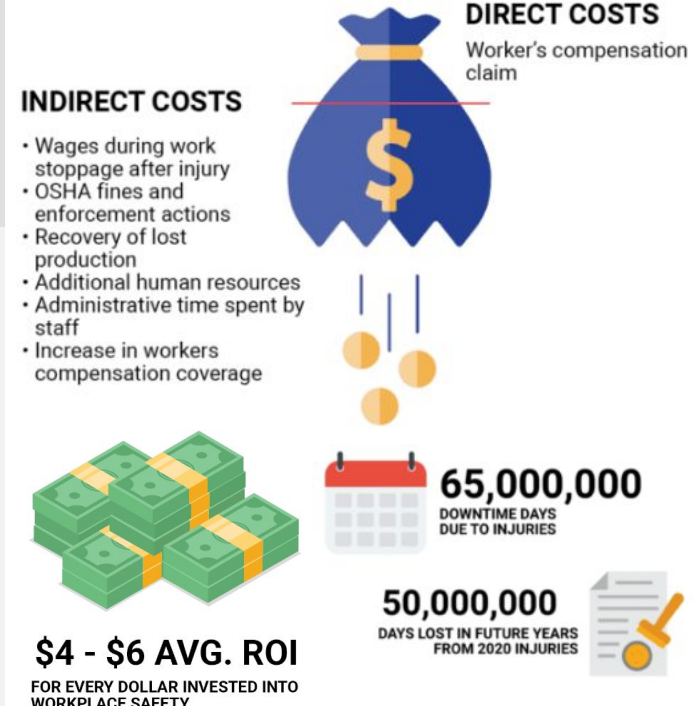
Risk Reduction: Enhanced compliance minimizes legal and financial risks and potential penalties

The maximum **penalties** will be raised from the current **$100,000** and/or up to one year imprisonment

Experience Enhancement: Effective incident response fosters a safer, positive work environment.

Organizations prioritizing safety lower safety incidents by **70%** according to a study by **Gallup 2017.**

The **NSC** estimated a total of **99,000,000 days lost** due to work-related injuries for 2020.



Solution Benefits with Examples of applications

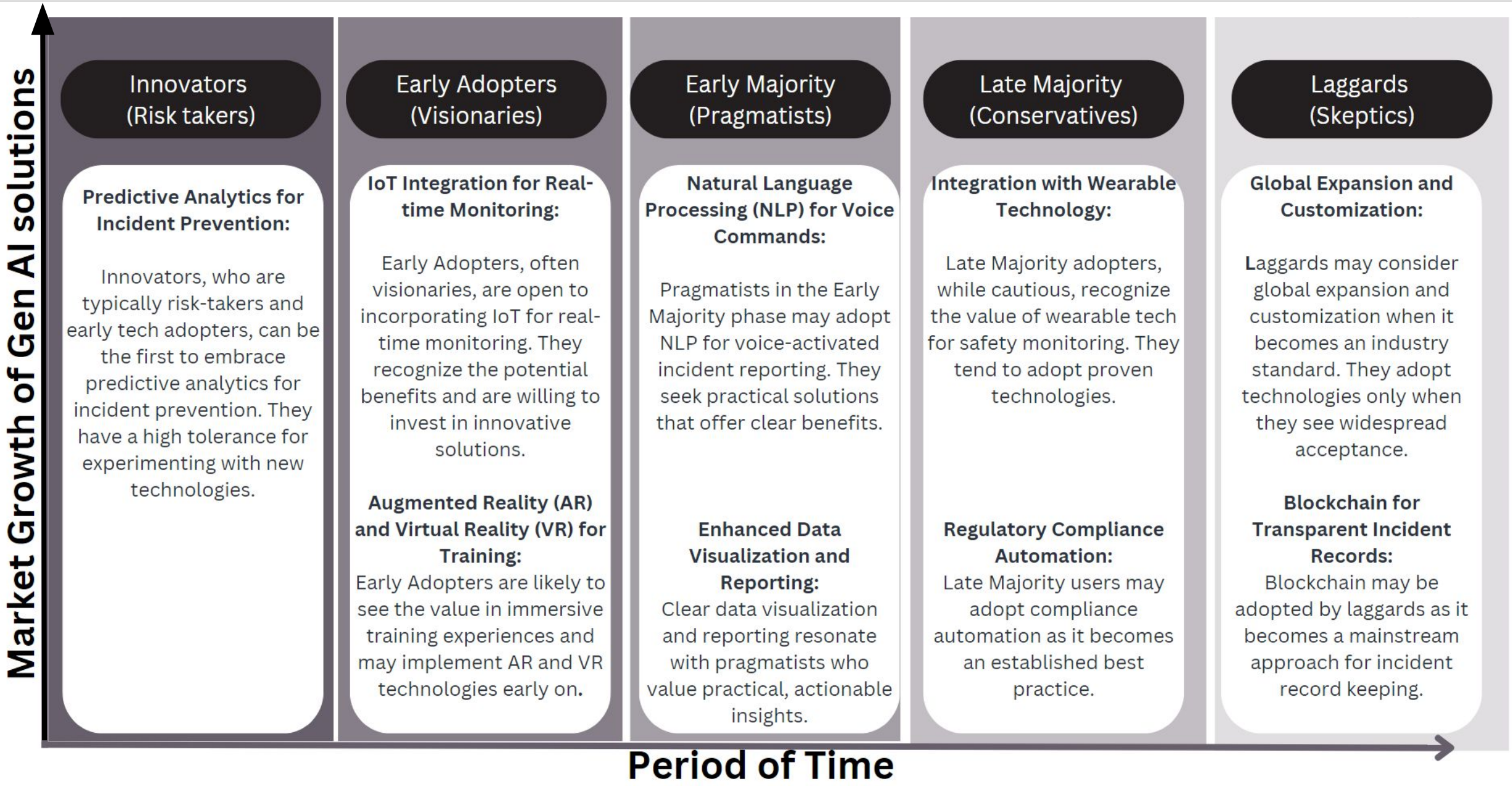
Solution	Efficient Incident Reporting:	Real-time Compliance Reporting	Faster Approval of Medical Insurance
Tangible Benefit	• Reduction in incident reporting time • Intangible Benefit: Enhanced situational awareness for quicker response.	• Reduction in compliance reporting time. • Improved regulatory adherence leading to a better organizational image.	• Medical insurance approval process shortened • Increased trust and satisfaction among employees.
Intangible Benefit			
Financial Benefit	• Financial Benefit: Potential cost savings due to reduced downtime and operational disruptions.	• Lower compliance-related fines and penalties.	• Reduced financial burden on the organization for medical expenses.

Solution	Root Cause Analysis	Automated Compliance Reporting	Improved Decision-making
Tangible Benefit	• Improvement in identifying root causes promptly. • More effective preventive measures reducing future incidents.	• Reduction in manual compliance reporting efforts. • Enhanced transparency and trust with regulatory bodies.	• Faster decision-making with access to real-time safety data. • Enhanced operational efficiency and risk management.
Intangible Benefit			
Financial Benefit	• Cost reduction through prevention of recurrent incidents and associated expenses.	• Lower compliance-related administrative costs.	• Improved profitability through better-informed strategic decisions.

Revenue Benefits & Facts supporting it

• **Reduced Downtime**: Minimizing incidents and improving response times can lead to significant revenue savings by reducing downtime in mining and energy operations.	• The average cost of lost revenue, financial penalties, idle staff time and restarting lines is $532,000 per hour, amounting to $172 million per plant annually.
• **Better Reputation:** Enhanced safety practices and compliance can attract new business opportunities and partnerships, contributing to revenue growth.	• 92% will be more likely to trust a company that supports social or environmental issues.
• **Insurance Cost Savings:** Efficient medical insurance approvals can result in lower insurance premiums, contributing to cost reduction.	• The digitization of the claims and filing process can improve the customer experience by 20% points and reduces expenses by 25-30%
• **Data-Driven Efficiency:** Improved decision-making can optimize resource allocation, reduce operational costs, and increase overall profitability.	• According to a report by McKinsey, companies that effectively use data analytics for decision-making can achieve up to a 20% improvement in operational efficiency.
• **Preventing Regulatory Fines:** Avoiding regulatory fines and penalties can protect revenue and maintain financial stability.	• According to a report by McKinsey, effective use of digital technologies in risk management can lead to a 10-25% reduction in regulatory costs in the mining industry.
• **Market Share Growth:** A better safety record can help capture a larger share of the market and increase revenue potential.	• According to OSHA, effective safety and health management systems can reduce workplace injuries and illnesses by 20-40%. This can directly impact turnover rates.
• **Reduced Employee Turnover:** Improved safety and faster medical attention can reduce employee turnover and recruitment costs.	

Future Scope of solution focusing on clients at different stages of Tech Adoption cycle



Product Demonstration Video and Repository Links

Link to Demo Prototype

Link to GitHub Repository

THANK YOU