

# Statement Of Work: Project Guffaw (LLM-Based Humor Generation)

This Statement of Work (SOW) defines the tasks, deliverables, and schedule for the collaboration between **Microsoft Corporation** ("Client") and **Craig Frint** ("Contractor").

## Statement of work template

### Title

Microsoft AI R&D: Project Guffaw (LLM-Based Humor Generation)

### Abstract

This project involves the research, development, and implementation of advanced prompt engineering and fine-tuning techniques for Large Language Models (LLMs) to generate high-quality, contextually relevant, and "hilarious" humor. Craig Frint will be responsible for creating a specialized humor engine that can adapt to various comedic styles (e.g., wordplay, situational irony, satire) while maintaining strict safety and toxicity filters. The primary output is a robust API integration and a fine-tuned model checkpoint capable of outperforming baseline GPT-4 results in human-rated "humor tests."

### Value

The estimated value of this work is **\$120,000 USD**. This includes all professional services provided by Craig Frint, covering research, dataset curation, model training, and integration. The budget is allocated across five milestones, with payments triggered by specific delivery results. Microsoft will provide all necessary Azure cloud compute resources and access to internal LLM APIs separately from this fee.

### Scope

The scope encompasses the end-to-end development of the humor engine within the Microsoft Azure AI ecosystem. Work includes:

1. **Dataset Curation:** Scraping and cleaning a high-quality "joke bank" of 50,000+ entries.
2. **Prompt Engineering:** Developing complex chain-of-thought prompts to improve comedic timing.
3. **Fine-tuning:** Executing LoRA/QLoRA fine-tuning on Microsoft-hosted base models.
4. **Evaluation Framework:** Designing a "Humor Scorecard" utilizing human-in-the-loop (HITL) feedback.
5. **Integration:** Building a RESTful API wrapper for internal Microsoft teams to consume.

## Type

This is a work-for-hire agreement as described by U.S. law. All intellectual property (IP), including code, training configurations, and curated datasets, created during this project by Craig Frint remains the exclusive property of Microsoft Corporation.

## Payment

The total budget of \$120,000 will be paid as follows:

- **Milestone 1 (Project Kickoff & Data Curation):** 15% (\$18,000)
- **Milestone 2 (Prototype Prompt Engine):** 20% (\$24,000)
- **Milestone 3 (Fine-tuned Model Checkpoint):** 25% (\$30,000)
- **Milestone 4 (API Integration & Testing):** 25% (\$30,000)
- **Milestone 5 (Final Documentation & Handover):** 15% (\$18,000)

Payments will be made via bank transfer within 30 days of milestone approval.

## Purpose

### Objectives

- Develop an LLM capable of generating humor across five distinct categories (Puns, Satire, Observational, Surreal, and Wordplay).
- Achieve a "Joke Success Rate" of >75% (defined as a 4/5 star rating from human evaluators).
- Implement a real-time safety layer to filter out offensive or non-inclusive content.

## Performance

Performance will be measured by:

- **Average Rating (AR):** Mean human rating (1-5 scale) for generated jokes.
- **Novelty Score:** Percentage of jokes not found in the training dataset (target >90%).
- **Latency:** API response time for a "single joke" request must be <800ms.
- **Toxicity Rate:** <0.01% pass-through of flagged content in stress tests.

## Factors

- **Azure Compute Availability:** Craig Frint requires consistent access to A100/H100 clusters for fine-tuning.
- **Human Eval Access:** Timely feedback from Microsoft's internal crowd-sourcing team is essential for iterative improvement.
- **API Stability:** The project assumes the stability of the Azure OpenAI base models.

## Who does what

## People

- **Craig Frint:** Lead Software Engineer (Contractor).
- **Dr. Satya Joke-wala:** Microsoft Project Sponsor / Work Authority.
- **Microsoft AI Red Team:** Safety and Ethics Auditors.
- **Internal UX Team:** End-user testers.

## Roles

- **Contractor:** Responsible for code, data processing, and model performance.
- **Work Authority:** Responsible for project sign-off, resource allocation, and strategic direction.

## Responsibilities

Area	Craig Frint	Microsoft
Model Architecture	R	A/C
Data Provision	C	R
Compute Infrastructure	I	R
Final Acceptance	C	A
(R=Responsible, A=Accountable, C=Consultable, I=Informable)		

## Context

### Past

Microsoft's previous attempts at "AI Humor" (Project Chuckle) resulted in models that were either too repetitive or inadvertently offensive. This project seeks to leverage New LLM architectures (GPT-4o/Llama-3 variants) to solve the "timing" and "irony" gaps identified in 2023.

### Present

Current consumer AI assistants are often perceived as "stiff" or "robotic." Integrating a specialized humor module is a key differentiator for the next generation of Windows Copilot features.

## Future

Success in Project Guffaw will lead to the "Humor-as-a-Service" (HaaS) API, intended to be used across Xbox, Windows, and LinkedIn for lighthearted user engagement.

## Planning

### Requirements

Craig Frint shall deliver:

- Python-based data pipeline for joke cleaning.
- A set of system prompts (GPT-compatible) optimized for humor.
- A fine-tuned Llama-3-70b (or similar) adapter.
- Technical documentation in Markdown format.

### Specifications

Code must adhere to Microsoft's internal Python style guide. APIs must use JSON for requests/responses and include headers for tracking "Joke Type" and "User Mood."

### Work breakdown structure (WBS)

1. **Phase 1:** Data collection & De-duplication (Weeks 1-4).
2. **Phase 2:** Prompt engineering & Baseline Testing (Weeks 5-8).
3. **Phase 3:** Fine-tuning & Weight Optimization (Weeks 9-16).
4. **Phase 4:** API Deployment & Red-teaming (Weeks 17-20).
5. **Phase 5:** Final Human Evaluation & Sign-off (Weeks 21-24).

### Method and source of acceptance

Acceptance is based on the "**The Comedy Club**" test: 100 jokes are presented to a blind panel of 10 human judges. If 75 jokes receive a "laugh/smile" rating, the milestone is accepted.

## Other terms and conditions

### Authorities

- **Work Authority:** Dr. Satya Joke-wala (General Manager, AI Humor Division).
- **Invoicing:** Microsoft Accounts Payable.

### Contractor's obligations

- Maintain strict confidentiality via Microsoft NDA.
- Provide weekly status reports via Microsoft Teams.
- Use internal GitHub repositories for all version control.

### Security requirements

Craig Frint must complete the "Security and Privacy for AI" training and use a Microsoft-issued laptop for all development involving internal data.

## Schedule

### Expected start date and completion date

- **Start:** April 1, 2024
- **End:** September 30, 2024

## Wordbook

- **LLM:** Large Language Model.
- **LoRA:** Low-Rank Adaptation (fine-tuning technique).
- **Prompt Inject:** A technique used to bypass safety filters (to be defended against).
- **Guffaw:** A loud and hearty laugh.

## Sign-off

NOTE: Before signing the Statement of Work, if you have any questions or concerns, please call the Work Authority indicated above to negotiate any issues.

### For Microsoft Corporation:

Printed Name: \_\_\_\_\_

Signature: \_\_\_\_\_

Date: \_\_\_\_\_

### Contractor:

Printed Name: **Craig Frint**

Signature: \_\_\_\_\_

Date: \_\_\_\_\_