

Project Summary: The Story Synthesiser (One-Day Hack Day)

Objective: To produce a summary document that provides the complete background for the "Made Impact" and "Million Stories" initiative. This document will detail the two-part plan for the upcoming one-day hack day, which involves 10 software engineering graduates. The goal is to create a proof-of-concept for (1) a synthetic "Story Generator" and (2) an "Editorial Workflow" tool.

1. Background: The "Made Impact" & "Million Stories" Initiative

The one-day project is in service of a much larger, long-term initiative called "Made Impact".

- **Founder:** Chris Arnold.
- **Core Mission:** The primary project is the "Million Stories" initiative. The goal is to collect and analyze one million stories of international exchange.
- **The "Why":** The core belief is that international travel and cultural exchange are a powerful force for world peace, fostering understanding and collaboration.
- **Ultimate Goal:** To use the data from these one million stories to advocate on behalf of the travel industry. The ambitious, long-term goal is to build a case for the travel industry to receive a Nobel Peace Prize.
- **Funding:** The project is currently bootstrapped by its founder. The long-term strategy is to partner with travel providers who would gain access to the data and insights as a form of corporate social responsibility (CSR) and marketing.
- **Story Collection Strategy:** Real stories will be collected via three main avenues:
 1. Public submissions on the website.
 2. A "world storyteller project".
 3. Studio days, filmmaker collaborations, and collecting stories at conferences.

2. The Problem & The Hack Day Solution

- **The Problem:** It will take a very long time to collect one million real stories. To build the analysis platform and understand the data, we need a large dataset *now*. This synthetic dataset will help test the data model, understand what's missing, and prepare for handling diverse content, including negative stories.
- **The Solution (Our Project):** The "Story Synthesiser". We will host a one-day hack day with 10 graduates to build a **proof of concept (PoC)**. The motto for the day is "**you're either winning or you're learning**"; success is defined by learning and proving the concept, not necessarily delivering a polished product. Even one successfully generated story is a "win".

3. The Core Data Model

The synthetic data must conform to the project's data model.

- **User / Profile:** A user has a name, home country, date of birth, gender and answers a survey question about their belief in international exchange as a force for peace (ranked 1 – 10)

- **Experience (One-to-Many):** A user can have multiple experiences.
 - **Data:** Captures destination country, type of experience (e.g., summer camp, Aupair), and provider.
 - **Survey:** Asks about improvements in soft skills (e.g., communication, resilience).
- **Story (One-to-Many):** An experience can have multiple stories, each capturing a specific *moment* or interaction.
 - **Content:** Can be text, video, or audio. The plan is to get AI transcripts from video/audio for analysis. Photos are also desired.
 - **Survey:** Asks about changes in *compassion*.

4. The Two-Part Hack Day Plan

The 10 graduates will be split into two teams, each tackling one part of the proof-of-concept.

Part 1: The Story Generator (Proof of Concept)

This team will build the tool that generates the synthetic data.

- **Objective:** Create a web-based tool that uses a local LLM to generate a batch of plausible, structured stories that match our data model.
- **Key Workflow & Features:**
 1. **Web UI for Parameters:** A simple UI (React or pure HTML/CSS for speed) where an internal user can set parameters.
 2. **Configurable Parameters:**
 - Number of users/stories to generate.
 - Percentage of outliers (e.g., 95% positive stories, 5% negative stories).
 - Simulated user attributes (e.g., literacy/writing skills).
 - A "seed" for noise, to ensure repeatable generation if possible.
 3. **Generation Process:** The tool will take these parameters and prompt a local LLM (e.g., Llama, Ollama). The process will:
 - Generate a fake user (with home country, skills).
 - Generate an experience for that user (destination, type).
 - Generate a story based on those details.
 - Repeat this N times.
 4. **Output:** The tool will output all the generated stories as a single structured data file (e.g., JSON or CSV) for download.

Part 2: The Editorial Workflow (Proof of Concept)

This team will build a tool to manage and utilize the stories generated in Part 1.

- **Objective:** Create a web-based "Story Browser" that uses AI to analyze the synthetic stories and explore an editorial workflow for publishing them.
- **Key Workflow & Features:**
 1. **AI Story Analysis:** (This assumes an output file from Part 1 is available). Use AI to analyze each generated story for key metrics:
 - Sentiment (positive, negative, heartwarming).

- Quality (well-written, poorly written).
 - Emotional Resonance.
 - Themes.
2. **"Story Browser" UI:** Build a simple web interface that displays the generated stories in a list or table, along with their new AI-generated metadata (Sentiment, Quality, etc.).
 3. **Editorial Curation:** The UI should allow an editor or marketer to "cherry-pick" the best or most interesting stories.
 4. **Draft Article Generation (Workflow):** The ultimate PoC goal. When an editor selects a story, the tool should trigger a workflow. This workflow will:
 - Take the chosen story.
 - Prompt an LLM with the story and the "Made Impact" brand voice guidelines.
 - Generate a **draft article** based on that story.
 - (Ideally) Show how this draft could be pushed to a CMS (like Sanity) for a "human-in-the-loop" to review, edit, and publish.