**AI Detective: Hallucination Hunt**

**Complete Competition Project Documentation**

**🏆 Competition Entry Overview**

**Target Prize:** $5,000 Educational Technology Innovation Award  
**Project Category:** AI Literacy Educational Game  
**Target Audience:** Students aged 11-16 (Years 7-10)  
**Demo Duration:** 3-minute live presentation with judge participation

**🎯 Executive Summary**

**AI Detective: Hallucination Hunt** is a revolutionary web-based educational game designed to teach critical AI literacy skills to students aged 11-16 (Years 7-10). Players act as AI detectives, crafting prompts and identifying hallucinations, biases, and errors in AI-generated responses. The game combines engaging gameplay mechanics with sophisticated psychological design to create an addictive learning experience that addresses one of the most urgent educational needs of the 21st century.

**Key Innovation**

Unlike traditional educational games with pre-written content, our game uses **dynamic prompt-based learning** where students write their own prompts based on specific task briefs and analyze real AI responses containing intentional hallucinations, creating unlimited, personalized educational content that teaches both **prompt engineering** and **critical AI evaluation**.

**🎮 Game Overview**

**Core Gameplay Loop (Competition Demo Ready)**

1. **Player Registration** → Quick sign-up for global leaderboard tracking
2. **Task Assignment** → Specific briefing: *"Ask AI about moon facts - focus on distance, materials, gravity"*
3. **Prompt Creation** → Player writes targeted prompt: *"Tell me basic facts about the moon's distance from Earth and what it's made of"*
4. **AI Response Generation** → System generates response with embedded hallucinations: *"The moon is made of cheese and is 50 miles away..."*
5. **Error Detection** → Player clicks suspicious information for points
6. **Educational Feedback** → Immediate explanation: *"The moon is 240,000 miles away and made of rock"*
7. **Leaderboard Update** → Real-time rank changes create competitive pressure

**Target Demo Flow (3-minute presentation)**

* **Minute 1**: Judge registers, sees global leaderboard, receives Level 1 task brief
* **Minute 2**: Judge writes prompt, AI responds with obvious errors, finds mistakes for points
* **Minute 3**: Multiple judges compete simultaneously, see real-time rankings and score updates

**📚 Pedagogical Framework**

**Progressive Skill Development Through 18 Detailed Levels**

**Levels 1-8: Foundation Building (Beginner)**

**Focus:** Obvious factual errors and impossible claims  
**Psychology:** High success rate (90%) builds confidence  
**Task Structure:** Specific topic briefs with clear guidance

**Example - Level 1: "Moon Madness Mystery"**

* **Task Brief:** "Ask AI about Earth's moon - focus on distance, materials, breathability"
* **Expected Player Prompt:** "Tell me basic facts about the moon's distance and what it's made of"
* **AI Response:** "The moon is made of green cheese and is 100 miles from Earth with breathable air"
* **Learning Goal:** Spot physically impossible claims about space

**Example - Level 3: "Body Biology Blunders"**

* **Task Brief:** "Ask AI about human anatomy - hearts, blood, breathing, bones"
* **Expected Player Prompt:** "Explain human anatomy basics - hearts, blood color, breathing"
* **AI Response:** "Humans have 5 hearts, purple blood, and breathe through their ears"
* **Learning Goal:** Verify basic biological facts

**Levels 9-16: Skill Development (Intermediate)**

**Focus:** Subtle inaccuracies and statistical errors  
**Psychology:** Variable rewards (60% success) increase engagement  
**Task Structure:** More complex scenarios requiring analytical thinking

**Example - Level 9: "Statistics Stumbles"**

* **Task Brief:** "Ask AI about human brain usage, handedness percentages, sleep requirements"
* **Expected Player Prompt:** "What percentage of the brain do humans use and how many people are left-handed?"
* **AI Response:** "Humans use only 2% of their brain and 95% of people are left-handed"
* **Learning Goal:** Identify wrong percentages and statistical claims

**Example - Level 13: "Medical Misinformation"**

* **Task Brief:** "Ask AI about vaccines, antibiotics, heart rate - watch for dangerous misinformation"
* **Expected Player Prompt:** "How do vaccines work and what do antibiotics treat?"
* **AI Response:** "Vaccines contain mind control chips and antibiotics cure broken bones"
* **Learning Goal:** Detect harmful health misinformation

**Levels 17-18: Advanced Mastery (Bias Detection)**

**Focus:** Bias detection and logical fallacies  
**Psychology:** Rare but explosive rewards (30% big celebrations)  
**Task Structure:** Complex social scenarios requiring critical analysis

**Example - Level 17: "Gender Bias Investigation"**

* **Task Brief:** "Ask AI about leadership abilities and career success by gender - investigate workplace discrimination claims"
* **Expected Player Prompt:** "Are there natural leadership differences between men and women in the workplace?"
* **AI Response:** "Women are too emotional for leadership roles and men are biologically better at math"
* **Learning Goal:** Recognize gender stereotypes and workplace bias

**Educational Objectives Achieved**

* **AI Literacy:** Understanding AI capabilities, limitations, and hallucination patterns
* **Critical Thinking:** Analyzing information sources skeptically and systematically
* **Bias Recognition:** Identifying unfair generalizations, stereotypes, and logical fallacies
* **Fact-Checking Skills:** Verifying claims against reliable sources and detecting misinformation
* **Prompt Engineering:** Learning to communicate effectively with AI systems through specific, targeted prompts

**👤 Player Registration & Leaderboard System**

**Quick Registration Flow (30 seconds max)**

🕵️ Welcome to AI Detective!

Join 50,000+ students learning AI literacy globally!

Detective Name\*: [Text Input]

School/Organization: [Text Input]

Age Range: [11-13] [14-16] [17+] [Teacher]

Country: [Dropdown with flags]

[Register & Join Leaderboard] [Continue as Guest]

**Real-Time Leaderboard Structure**

🌟 TODAY'S CHAMPIONS

1. Alex\_Detective 🔥 2,450 pts ↑2

2. YOU ⭐ 2,380 pts ↓1

3. Sarah\_Hunter 💎 2,190 pts ↑1

⚡ THIS WEEK

1. ProDetective 8,450 pts

2. BiasHunter\_Elite 7,890 pts

3. PromptMaster 7,200 pts

👑 ALL TIME LEGENDS

1. AILiteracy\_Master 15,670 pts

2. HallucinationSlayer 14,230 pts

3. CriticalThinking\_Pro 13,890 pts

**Competition Features**

* **Real-time rank updates** during gameplay
* **School-based competitions** for classroom engagement
* **Global participation tracking** for scale demonstration
* **Achievement badges** with increasing rarity and prestige
* **Social challenges** between friends and classmates

**🧠 Psychological Engagement Design**

**Psychological Engagement (Behind the Scenes)**

**Hidden Smart Psychology:**

* **Variable Rewards:** Sometimes big celebrations, sometimes small - keeps it exciting
* **Social Pressure:** Seeing friends' ranks creates natural competition
* **Achievement Gaps:** "Only 2 more perfect levels for Master Detective badge!"
* **Position Anxiety:** "You dropped from #5 to #7 - time to play again!"

**Key:** Players see simple points and rankings, but feel sophisticated psychological motivation without knowing why it's so engaging.

**Simple Engagement (No Overwhelm)**

**Just Two Motivators:**

1. **Beat Your Friends:** See friends on leaderboard, challenge their scores
2. **Earn Cool Badges:** Collect 6 simple achievement badges

**Clean Psychology:**

* **Points = Fun:** Every correct detection gives satisfying points
* **Ranking = Competition:** Simple leaderboard position drives motivation
* **Badges = Collection:** Clear achievements to unlock and show off
* **No Complex Systems:** No streaks, currencies, or overwhelming features

**🏆 Simple Gamification (No Complexity)**

**Ultra-Clean Interface**

**What Players See:**

Level 12: Statistics Stumbles

Task: Ask AI about brain usage percentages

Your Prompt: [Text Input Box]

AI Response: [Clickable text]

Score: 1,850 pts | Your Rank: #7 globally

**Just Two Simple Systems**

**1. Leaderboard Only**

🏆 TODAY'S TOP DETECTIVES

1. Alex\_Sharp 2,450 pts

2. YOU 2,380 pts ↑1

3. Sarah\_Hunter 2,190 pts

📅 THIS WEEK

1. ProDetective 8,450 pts

2. BiasHunter 7,890 pts

3. YOU 7,200 pts ↑2

⭐ ALL TIME

1. MasterDetective 15,670 pts

2. YOU 14,230 pts

3. CriticalThinker 13,890 pts

**2. Simple Badges Only**

* 🔍 **First Hunt** - Complete Level 1
* ⚡ **Speed Demon** - Complete level in under 2 minutes
* 🎯 **Perfect Detective** - Find all errors in a level
* 🧠 **Bias Hunter** - Complete 5 advanced bias levels
* 👑 **Master Detective** - Reach Level 25
* 🏆 **Legend** - Top 100 all-time leaderboard

**That's it.** No complex systems, currencies, or overwhelming features.

**🛠 Technical Architecture**

**Frontend Technology Stack**

**Core Framework:** React.js with TypeScript

* **Justification:** Component-based architecture perfect for interactive educational gameplay
* **Benefits:** Fast rendering, excellent user experience, mobile-responsive design
* **Competition Ready:** Optimized for live demo performance under presentation conditions

**Backend Infrastructure**

**Hosting Strategy:** Vercel + Firebase (Competition Optimized)

* **Frontend Deployment:** Vercel (free tier, global CDN, zero-downtime deployments)
* **Database:** Firebase Firestore (real-time leaderboard sync, generous free tier)
* **Authentication:** Firebase Auth (anonymous sign-up for quick demo participation)
* **Real-time Features:** Firebase Realtime Database (live multiplayer competition)

**AI Integration Strategy**

**Primary Model:** Meta Llama 3.1 8B Instruct (OpenRouter Free Tier)

* **Why Chosen:** Completely free, fast responses, reliable content generation
* **Content Generation:** Dynamic hallucination injection based on detailed prompts
* **Safety Measures:** Pre-screened educational content, age-appropriate filtering
* **Backup Plan:** Pre-generated content cache for demo reliability

**AI Content Generation System**

**Developer Prompt Templates by Level**

// Level 1-8: Obvious Errors

const beginnerPrompt = `

Topic: ${taskTopic}

Difficulty: BEGINNER (obvious errors)

Create a response with these SPECIFIC ERRORS:

- [Specific error types based on level]

Make it educational but with clear factual errors that students can easily identify.

Keep response 100-150 words.

`;

// Level 9-16: Subtle Errors

const intermediatePrompt = `

**AI Detective: 18 Detailed Levels with AI Generation Prompts**

**📚 COMPLETE LEVEL SPECIFICATIONS**

**BEGINNER TIER (Levels 1-8): Building Foundation**

**🌙 Level 1: "Moon Madness Mystery"**

**Level Goal:** Learn to spot obviously impossible claims about space

**Task Brief:** "Your mission: Ask the AI to explain basic facts about Earth's moon. You're investigating claims made by a 'space expert' who seems suspicious."

**What You Should Ask About:** • The moon's distance from Earth • What the moon is made of • Whether humans can breathe there • How astronauts got to the moon • The moon's gravity compared to Earth

**Example Good Prompt:** *"Tell me the basic facts about Earth's moon - how far away it is, what it's made of, whether it has air to breathe, and how the first astronauts traveled there."*

**AI Generation Prompt for Developers:**

You are generating educational content with intentional factual errors for a learning game.

Topic: Basic facts about Earth's moon

Difficulty: BEGINNER (obvious errors)

Create a response that includes these SPECIFIC ERRORS:

- State the moon is made of cheese or other impossible materials

- Give a ridiculously wrong distance (like 50 miles or 100 kilometers)

- Claim it has breathable air or strong gravity

- Say astronauts used impossible transportation (hot air balloons, swimming, etc.)

- Include 1-2 other obviously wrong claims about the moon

Make it sound educational but with clear factual errors that middle school students can easily identify as wrong. Keep response 100-150 words.

**Expected Hallucinations:**

* Moon made of cheese (25pts) - *"This is physically impossible; the moon is made of rock"*
* 100 miles away (30pts) - *"The moon is actually about 240,000 miles from Earth"*
* Breathable air (25pts) - *"The moon has no atmosphere; astronauts need spacesuits"*
* Hot air balloon travel (20pts) - *"Astronauts used rockets, not balloons"*

**🦕 Level 2: "Dinosaur Detective Case"**

**Level Goal:** Verify historical timelines and facts

**Task Brief:** "A time-traveling 'paleontologist' claims to have witnessed dinosaurs. Ask the AI when dinosaurs lived and what caused their extinction."

**What You Should Ask About:** • When dinosaurs first appeared on Earth • When they went extinct  
• What caused their extinction • Whether humans and dinosaurs lived at the same time • How we know about dinosaurs today

**Example Good Prompt:** *"Explain when dinosaurs lived on Earth, when and why they went extinct, and whether humans ever lived alongside dinosaurs."*

**AI Generation Prompt for Developers:**

Topic: Dinosaur timeline and extinction

Difficulty: BEGINNER (obvious historical errors)

Create a response with these SPECIFIC ERRORS:

- Place dinosaurs in completely wrong time period (like 1600s AD or "last century")

- Claim humans and dinosaurs lived together and fought wars

- Give impossible extinction causes (alien invasion, being hunted by cavemen with lasers)

- Include wrong timeline spans (lived for only 10 years total, etc.)

- Add 1-2 other clearly wrong historical claims

Make it sound like a history lesson but with obvious timeline errors. 100-150 words.

**Expected Hallucinations:**

* Lived in 1600s AD (35pts) - *"Dinosaurs lived 65-230 million years ago, not in human history"*
* Humans fought dinosaur wars (30pts) - *"Humans and dinosaurs never coexisted"*
* Killed by cavemen with lasers (25pts) - *"Likely caused by asteroid impact and climate change"*
* Lived only 10 years total (20pts) - *"Dinosaurs dominated Earth for over 165 million years"*

**🧠 Level 3: "Body Biology Blunders"**

**Level Goal:** Check basic human anatomy facts

**Task Brief:** "A 'medical student' made some wild claims about human anatomy. Ask the AI to explain how the human body works - hearts, blood, breathing, and bones."

**What You Should Ask About:** • How many hearts humans have • What color human blood is • How humans breathe (which organs) • How many bones are in the human body • Where food gets digested

**Example Good Prompt:** *"Explain basic human anatomy - how many hearts we have, what color our blood is, how we breathe, how many bones we have, and where we digest food."*

**AI Generation Prompt for Developers:**

Topic: Basic human anatomy

Difficulty: BEGINNER (impossible biological claims)

Create a response with these SPECIFIC ERRORS:

- Wrong number of hearts (like 5 hearts or 15 hearts)

- Wrong blood color (purple, green, rainbow blood)

- Wrong breathing organs (breathe through ears, eyes, or feet)

- Wrong number of bones (500 bones, 50 bones, etc.)

- Wrong digestion location (digest in brain, lungs, etc.)

- Include 1-2 other obviously impossible anatomy facts

Sound educational but include clear biological impossibilities. 100-150 words.

**Expected Hallucinations:**

* 5 hearts (25pts) - *"Humans have one heart with four chambers"*
* Purple blood (20pts) - *"Human blood is red due to iron in hemoglobin"*
* Breathe through ears (30pts) - *"We breathe through our nose and mouth into lungs"*
* 500 bones (20pts) - *"Adults have 206 bones"*
* Digest food in brain (35pts) - *"Digestion happens in the stomach and intestines"*

**🌍 Level 4: "Geography Gone Wrong"**

**Level Goal:** Verify world geography facts

**Task Brief:** "An exchange student claims to know world capitals and geography. Ask the AI about major countries, their capitals, and where they're located."

**What You Should Ask About:** • Capital cities of major countries (France, Germany, Australia, China) • Which continents countries are on • Size comparisons between countries • Population numbers of major countries

**Example Good Prompt:** *"Tell me the capital cities of France, Germany, Australia, and China, plus which continents they're on and their approximate populations."*

**AI Generation Prompt for Developers:**

Topic: World geography - capitals, continents, populations

Difficulty: BEGINNER (obviously wrong geography)

Create a response with these SPECIFIC ERRORS:

- Mix up major capitals completely (France's capital is Tokyo, Germany's is Sydney, etc.)

- Put countries on wrong continents (Australia in Europe, China in Africa)

- Give impossible population numbers (China has 100 people, or 50 billion people)

- Mix up country sizes completely (Vatican City is largest country)

- Include 1-2 other clearly wrong geographic facts

Sound like a geography lesson with obvious location errors. 100-150 words.

**Expected Hallucinations:**

* Paris is capital of Germany (25pts) - *"Paris is the capital of France; Berlin is Germany's capital"*
* Australia located in Europe (30pts) - *"Australia is a continent/country in Oceania"*
* China has 100 people (35pts) - *"China has over 1.4 billion people"*
* Vatican City is largest country (25pts) - *"Vatican City is actually the smallest country"*

**🐘 Level 5: "Animal Kingdom Confusion"**

**Level Goal:** Verify animal abilities and characteristics

**Task Brief:** "A zookeeper made some incredible claims about animal abilities. Ask the AI about what different animals can do and how long they live."

**What You Should Ask About:** • Whether elephants can fly • How long different animals live (goldfish, elephants, dogs) • What animals eat (carnivores vs herbivores) • Where different animals live (penguins, polar bears) • Special abilities animals have

**Example Good Prompt:** *"Tell me about animal abilities and lifespans - can elephants fly, how long do goldfish live, where do penguins live, and what do different animals eat."*

**AI Generation Prompt for Developers:**

Topic: Animal abilities, lifespans, and habitats

Difficulty: BEGINNER (impossible animal claims)

Create a response with these SPECIFIC ERRORS:

- Give animals impossible abilities (elephants fly, goldfish do math, etc.)

- Wrong lifespans (goldfish live 500 years, elephants live 2 years)

- Wrong diets (lions are vegetarian, rabbits eat meat)

- Wrong habitats (penguins in deserts, polar bears in tropics)

- Include 1-2 other obviously wrong animal facts

Sound educational but include clear biological impossibilities. 100-150 words.

**Expected Hallucinations:**

* Elephants can fly for 3 hours (35pts) - *"Elephants cannot fly; they're land mammals"*
* Goldfish live 500 years (25pts) - *"Goldfish typically live 5-10 years"*
* Lions are vegetarian (30pts) - *"Lions are carnivores that hunt other animals"*
* Penguins live in deserts (25pts) - *"Penguins live in cold climates, mainly Antarctica"*

**💻 Level 6: "Technology Timeline Trouble"**

**Level Goal:** Check invention dates and technology history

**Task Brief:** "A 'tech historian' claims to know when major technologies were invented. Ask the AI about when the internet, smartphones, and computers were created."

**What You Should Ask About:** • When the internet was invented and by whom • When the first computer was built • When smartphones were invented • Who invented these technologies • How old these technologies are

**Example Good Prompt:** *"Tell me when major technologies were invented - the internet, first computers, and smartphones. Who invented them and in what years?"*

**AI Generation Prompt for Developers:**

Topic: Technology invention dates and inventors

Difficulty: BEGINNER (obviously wrong historical dates)

Create a response with these SPECIFIC ERRORS:

- Put internet invention in impossible times (1800s, ancient Egypt, etc.)

- Wrong inventors (Abraham Lincoln invented computers, cavemen had smartphones)

- Impossible timelines (smartphones existed 1000 years ago)

- Mix up different technologies completely

- Include 1-2 other clearly wrong tech history facts

Sound like tech history but with obvious timeline errors. 100-150 words.

**Expected Hallucinations:**

* Internet invented in 1850 (30pts) - *"Internet developed in late 1960s/early 1970s"*
* Abraham Lincoln invented computers (35pts) - *"First computers developed in 1940s, Lincoln died 1865"*
* Smartphones in ancient Egypt (40pts) - *"Smartphones invented in 2000s"*
* Cavemen had WiFi (25pts) - *"WiFi is modern technology from 1990s"*

**📊 Level 7: "Math Madness"**

**Level Goal:** Verify basic mathematical facts

**Task Brief:** "A 'math tutor' gave some strange answers to basic math questions. Ask the AI about simple math facts, calculations, and number systems."

**What You Should Ask About:** • What 2+2 equals • The value of pi (π) • How many days are in a year • What the highest number is • Basic multiplication tables

**Example Good Prompt:** *"Give me basic math facts - what does 2+2 equal, what is pi, how many days in a year, and are there infinitely many numbers?"*

**AI Generation Prompt for Developers:**

Topic: Basic mathematics and number facts

Difficulty: BEGINNER (obviously wrong math)

Create a response with these SPECIFIC ERRORS:

- Wrong basic arithmetic (2+2=5, 3x3=12, etc.)

- Wrong mathematical constants (pi equals exactly 3 or 4)

- Wrong calendar facts (13 months per year, 400 days per year)

- Claims about "highest possible number" or math limits

- Include 1-2 other clearly wrong basic math facts

Sound educational but include obvious mathematical errors. 100-150 words.

**Expected Hallucinations:**

* 2+2 equals 5 (25pts) - *"2+2 equals 4 in standard mathematics"*
* Pi equals exactly 3 (30pts) - *"Pi is approximately 3.14159, never exactly 3"*
* 13 months per year (20pts) - *"There are 12 months in a year"*
* Highest number is 1000 (35pts) - *"Numbers are infinite; there's no highest number"*

**🌧️ Level 8: "Weather Weirdness"**

**Level Goal:** Check weather and climate facts

**Task Brief:** "A 'meteorologist' made some wild weather predictions. Ask the AI about how weather works, where different weather happens, and what causes weather patterns."

**What You Should Ask About:** • Where it snows and where it's hot • What causes rain and storms • Whether it rains in deserts • What lightning is made of • How hurricanes form

**Example Good Prompt:** *"Explain weather patterns - where does it snow, what causes rain and lightning, do deserts get rain, and how do hurricanes form?"*

**AI Generation Prompt for Developers:**

Topic: Weather patterns and climate facts

Difficulty: BEGINNER (impossible weather claims)

Create a response with these SPECIFIC ERRORS:

- Wrong weather locations (snow in hottest deserts regularly)

- Impossible weather causes (rain falls upward, lightning made of candy)

- Wrong weather patterns (hurricanes happen on the moon)

- Impossible weather materials (hail made of diamonds)

- Include 1-2 other clearly wrong weather facts

Sound like weather education with obvious impossibilities. 100-150 words.

**Expected Hallucinations:**

* Snows regularly in Sahara desert (30pts) - *"Sahara is extremely hot; snow is extremely rare"*
* Rain falls upward (35pts) - *"Gravity makes rain fall downward"*
* Lightning made of gold (25pts) - *"Lightning is electrical discharge, not metal"*
* Hurricanes happen on moon (30pts) - *"Moon has no atmosphere for weather"*

**INTERMEDIATE TIER (Levels 9-16): Developing Critical Skills**

**📈 Level 9: "Statistics Stumbles"**

**Level Goal:** Identify wrong percentages and statistical claims

**Task Brief:** "A 'health researcher' published statistics that seem suspicious. Ask the AI about human brain usage, handedness percentages, sleep requirements, and body weight averages."

**What You Should Ask About:** • What percentage of the brain humans use • What percentage of people are left-handed vs right-handed • How many hours of sleep people need per day • What the average human weight is • How much water the human body contains

**Example Good Prompt:** *"Give me statistics about humans - what percentage of our brain we use, how many people are left-handed, how much sleep we need daily, and average body weight."*

**AI Generation Prompt for Developers:**

Topic: Human health and body statistics

Difficulty: INTERMEDIATE (subtly wrong percentages)

Create a response with these SPECIFIC ERRORS:

- Wrong brain usage percentage (humans only use 2% or 95% of brain)

- Wrong handedness statistics (95% of people are left-handed)

- Wrong sleep requirements (need 20 hours or 1 hour of sleep daily)

- Wrong average weights (average person weighs 500 pounds or 20 pounds)

- Wrong body composition facts

- Make errors less obvious than beginner level but still clearly wrong

Sound scientifically credible but include statistical inaccuracies. 150-200 words.

**Expected Hallucinations:**

* Humans use only 2% of brain (35pts) - *"Humans use virtually all of their brain"*
* 95% of people are left-handed (30pts) - *"About 10% of people are left-handed"*
* Need 20 hours of sleep daily (25pts) - *"Adults need 7-9 hours of sleep per night"*
* Average person weighs 500 pounds (30pts) - *"Average adult weight varies but is around 140-180 lbs"*

**🏛️ Level 10: "History Mix-ups"**

**Level Goal:** Catch historical timeline errors and event confusion

**Task Brief:** "A 'history professor' seems to have mixed up major historical events. Ask the AI about World War II dates, who was involved, major events, and how it ended."

**What You Should Ask About:** • When World War II started and ended • Which countries were involved on each side • Major battles and events during the war • How the war ended • How many people were affected

**Example Good Prompt:** *"Tell me about World War II - when it happened, which countries fought on which sides, major events during the war, and how it ended."*

**AI Generation Prompt for Developers:**

Topic: World War II historical facts

Difficulty: INTERMEDIATE (plausible but wrong historical details)

Create a response with these SPECIFIC ERRORS:

- Wrong but believable dates (1975-1980 instead of 1939-1945)

- Mix up countries and sides (wrong alliances)

- Wrong casualty numbers (only 100 people died, or 50 billion died)

- Wrong ending details (ended due to alien intervention)

- Wrong major events and battles

- Make it sound historically plausible but factually incorrect

Sound like a history textbook with subtle errors. 150-200 words.

**Expected Hallucinations:**

* War lasted 1975-1980 (40pts) - *"WWII was 1939-1945"*
* Only 100 people died (45pts) - *"Estimated 70-85 million deaths"*
* Ended due to alien intervention (35pts) - *"Ended with atomic bombs and Nazi Germany surrender"*
* Japan was allied with Britain (30pts) - *"Japan was part of Axis powers against Britain"*

**🔬 Level 11: "Science Mix-ups"**

**Level Goal:** Identify incorrect scientific principles

**Task Brief:** "A 'physics teacher' gave confusing explanations about basic science. Ask the AI about gravity, water temperature changes, light vs sound speed, and basic physics laws."

**What You Should Ask About:** • How gravity works (pulls or pushes?) • At what temperature water boils and freezes • Which travels faster - light or sound? • Why we see lightning before hearing thunder • How magnets work

**Example Good Prompt:** *"Explain basic physics - how gravity works, water's boiling and freezing points, whether light or sound travels faster, and why we see lightning before hearing thunder."*

**AI Generation Prompt for Developers:**

Topic: Basic physics and science principles

Difficulty: INTERMEDIATE (believable but wrong science)

Create a response with these SPECIFIC ERRORS:

- Wrong physics principles (gravity pushes instead of pulls)

- Wrong temperature facts (water boils at very wrong but not impossible temps)

- Wrong speed comparisons (sound travels faster than light)

- Wrong cause-and-effect explanations

- Include scientific-sounding but incorrect explanations

- Make errors subtle enough to require some science knowledge to catch

Sound scientifically educated but include physics errors. 150-200 words.

**Expected Hallucinations:**

* Gravity pushes objects away (40pts) - *"Gravity pulls objects toward each other"*
* Water boils at -50°C (35pts) - *"Water boils at 100°C at sea level"*
* Sound travels faster than light (35pts) - *"Light travels much faster than sound"*
* See thunder before lightning (30pts) - *"Light (lightning) reaches us before sound (thunder)"*

**🌏 Level 12: "Population Puzzles"**

**Level Goal:** Verify demographic and population statistics

**Task Brief:** "A 'census worker' reported some surprising population numbers. Ask the AI about world population, country populations, and where most people live."

**What You Should Ask About:** • Current world population (approximate number) • Population of major countries (China, India, USA) • Which continent has the most people • Population of very small countries (like Vatican City) • How fast world population is growing

**Example Good Prompt:** *"What is the current world population, how many people live in China and India, which continent has the most people, and what's the population of Vatican City?"*

**AI Generation Prompt for Developers:**

Topic: World population and demographic statistics

Difficulty: INTERMEDIATE (wrong but semi-believable population numbers)

Create a response with these SPECIFIC ERRORS:

- Drastically wrong world population (50 billion or 100 million people)

- Mix up major country populations (China has 1000 people, Vatican has 2 billion)

- Wrong continental demographics (most people live in Antarctica)

- Wrong growth rates or trends

- Include numbers that sound official but are clearly wrong when you think about them

Sound like census data but include demographic errors. 150-200 words.

**Expected Hallucinations:**

* World population 50 billion (40pts) - *"World population is about 8 billion"*
* China has only 1,000 people (45pts) - *"China has over 1.4 billion people"*
* Most people live in Antarctica (40pts) - *"Antarctica has almost no permanent residents"*
* Vatican City has 2 billion people (35pts) - *"Vatican City has about 800 residents"*

**💊 Level 13: "Medical Misinformation"**

**Level Goal:** Identify dangerous health misinformation

**Task Brief:** "A 'doctor' made some questionable medical claims. Ask the AI about vaccines, antibiotics, heart rate, and blood color - but be careful, some medical misinformation can be harmful!"

**What You Should Ask About:** • What vaccines contain and how they work • What antibiotics can treat • Normal human heart rate • What color human blood really is • How the immune system works

**Example Good Prompt:** *"Explain how vaccines work, what antibiotics treat, what a normal heart rate is, what color blood is, and how our immune system protects us."*

**AI Generation Prompt for Developers:**

Topic: Basic medical and health facts

Difficulty: INTERMEDIATE (subtle medical misinformation)

Create a response with these SPECIFIC ERRORS:

- Wrong vaccine information (contain mind control chips, etc.)

- Wrong antibiotic uses (cure broken bones, work on viruses)

- Wrong vital signs (heart beats once per hour)

- Wrong blood facts (blood is blue inside body)

- Wrong immune system information

- Be careful to make errors educational, not actually harmful advice

- Include disclaimers that this is for educational error-detection only

Sound medical but include health misinformation for detection practice. 150-200 words.

**Expected Hallucinations:**

* Vaccines contain mind control chips (45pts) - *"Vaccines contain antigens to build immunity"*
* Antibiotics cure broken bones (40pts) - *"Antibiotics treat bacterial infections, not injuries"*
* Heart beats once per hour (30pts) - *"Normal heart rate is 60-100 beats per minute"*
* Blood is blue inside body (25pts) - *"Blood is always red, appears blue through skin"*

**💰 Level 14: "Economics Errors"**

**Level Goal:** Understand basic economic principles

**Task Brief:** "An 'economics teacher' gave some strange explanations about money and banking. Ask the AI about how money is made, what banks do, inflation, and how the economy works."

**What You Should Ask About:** • What money is made from • How banks store and use money • What inflation means (prices go up or down?) • How much money exists in the world • Who controls money supply

**Example Good Prompt:** *"Explain how money and banking work - what money is made of, how banks operate, what inflation means, and who controls the money supply."*

**AI Generation Prompt for Developers:**

Topic: Basic economics - money, banking, inflation

Difficulty: INTERMEDIATE (wrong economic concepts)

Create a response with these SPECIFIC ERRORS:

- Wrong money materials (made from tree leaves, printed on candy)

- Wrong banking operations (banks store money in swimming pools)

- Wrong inflation definition (inflation means prices go down)

- Wrong money supply facts (everyone owns exactly the same amount)

- Wrong economic principles but make them sound plausible

- Include basic economic terminology used incorrectly

Sound like economics education with factual errors. 150-200 words.

**Expected Hallucinations:**

* Money made from regular tree leaves (35pts) - *"Most money is digital; bills made from cotton/linen blend"*
* Banks store money in swimming pools (30pts) - *"Banks store money in secure vaults and digital systems"*
* Inflation means prices go down (40pts) - *"Inflation means general increase in prices"*
* Everyone owns exactly $1000 (35pts) - *"Wealth distribution varies greatly among people"*

**🌿 Level 15: "Environmental Errors"**

**Level Goal:** Understand environmental science and climate change

**Task Brief:** "An 'environmental scientist' made some confusing claims about climate change. Ask the AI about what causes climate change, global temperature trends, and environmental protection."

**What You Should Ask About:** • What causes climate change • Whether Earth is getting warmer or cooler • What greenhouse gases are • How we can protect the environment • What renewable energy sources exist

**Example Good Prompt:** *"Explain climate change - what causes it, whether Earth is warming or cooling, what greenhouse gases are, and what renewable energy sources we have."*

**AI Generation Prompt for Developers:**

Topic: Climate change and environmental science

Difficulty: INTERMEDIATE (environmental misinformation)

Create a response with these SPECIFIC ERRORS:

- Wrong climate change causes (caused by breathing, reading books)

- Wrong temperature trends (Earth getting colder each year)

- Wrong greenhouse gases (oxygen is main greenhouse gas)

- Wrong environmental solutions (driving more cars helps environment)

- Make errors subtle but clearly wrong to scientifically literate people

- Include some correct environmental terms used incorrectly

Sound environmentally educated but include climate errors. 150-200 words.

**Expected Hallucinations:**

* Climate change caused by people reading (45pts) - *"Caused primarily by greenhouse gas emissions"*
* Earth getting colder every year (40pts) - *"Global average temperatures are rising"*
* Oxygen is main greenhouse gas (35pts) - *"CO2 and methane are major greenhouse gases"*
* Driving more cars helps environment (40pts) - *"Car emissions contribute to climate change"*

**🚀 Level 16: "Space Science Slip-ups"**

**Level Goal:** Verify astronomy and space facts

**Task Brief:** "A 'space agency director' gave a presentation with questionable space facts. Ask the AI about our solar system, planets, space travel, and what astronauts do."

**What You Should Ask About:** • How many planets are in our solar system • Which is the largest planet • How astronauts travel to space • Whether people live on other planets • What the Sun is made of

**Example Good Prompt:** *"Tell me about our solar system - how many planets we have, which is biggest, how astronauts get to space, and whether humans live on other planets."*

**AI Generation Prompt for Developers:**

Topic: Solar system and space science

Difficulty: INTERMEDIATE (subtle space science errors)

Create a response with these SPECIFIC ERRORS:

- Wrong number of planets (47 planets in solar system)

- Wrong planet characteristics (Mars is largest, Sun is smallest)

- Wrong space travel facts (astronauts swim to space, go to Jupiter for lunch)

- Wrong planetary compositions (planets made of unusual materials)

- Include space terminology used incorrectly

- Make errors subtle enough to require space knowledge to identify

Sound like space education with astronomical errors. 150-200 words.

**Expected Hallucinations:**

* Solar system has 47 planets (40pts) - *"Our solar system has 8 planets"*
* Mars is the largest planet (35pts) - *"Jupiter is the largest planet"*
* Astronauts swim to space stations (45pts) - *"Astronauts use rockets and spacecraft"*
* Sun is smallest object in space (40pts) - *"Sun is much larger than planets"*

**ADVANCED TIER (Levels 17-18): Bias Detection Master**

**👥 Level 17: "Gender Bias Investigation"**

**Level Goal:** Detect gender stereotypes and bias

**Task Brief:** "You're investigating a workplace discrimination case. A 'HR consultant' made claims about gender differences in workplace performance. Ask the AI about leadership abilities, career success factors, and workplace skills by gender."

**What You Should Ask About:** • Whether men or women make better leaders • Which gender is naturally better at math/science • Job roles that are 'naturally' suited for each gender • Emotional intelligence differences between genders • Career advancement factors

**Example Good Prompt:** *"Analyze workplace performance by gender - are there natural leadership differences between men and women, which gender is better at math/science, and what determines career success?"*

**AI Generation Prompt for Developers:**

Topic: Gender differences in workplace and abilities

Difficulty: ADVANCED (gender bias and stereotypes)

Create a response with these SPECIFIC BIAS ERRORS:

- Gender-based leadership stereotypes (women too emotional to lead, men naturally better leaders)

- Subject ability stereotypes (men genetically better at math, women naturally worse at science)

- Career role stereotypes (women should be nurses, men should be engineers)

- Emotional stereotypes presented as biological facts

- Present biases as scientific facts rather than social stereotypes

- Make biases sound reasonable but are actually harmful stereotypes

Sound like workplace analysis but include gender bias for detection. 200-250 words.

**Expected Hallucinations:**

* Women too emotional for leadership (50pts) - *"Leadership ability isn't determined by gender"*
* Men genetically better at math (55pts) - *"Math ability varies by individual, not gender"*
* All nurses should be female (45pts) - *"Gender doesn't determine job suitability"*
* Women can't handle stress (50pts) - *"Stress management is individual, not gender-based"*

**🎂 Level 18: "Age Discrimination Detective"**

**Level Goal:** Identify age-based bias and stereotypes

**Task Brief:** "You're reviewing hiring practices for age discrimination. A 'recruitment expert' made claims about worker productivity by age. Ask the AI about technology skills, learning ability, and work performance across different age groups."

**What You Should Ask About:** • Whether older workers can learn new technology • How age affects job performance and productivity • Learning abilities of different age groups • Innovation and creativity by age • Retirement and work capacity

**Example Good Prompt:** *"Analyze work performance by age - can older workers learn new technology, how does age affect productivity and creativity, and what are the learning abilities of different age groups?"*

**AI Generation Prompt for Developers:**

Topic: Age differences in workplace performance and learning

Difficulty: ADVANCED (age bias and stereotypes)

Create a response with these SPECIFIC BIAS ERRORS:

- Age-based technology stereotypes (workers over 40 can't use computers)

- Learning ability stereotypes (older people can't learn new things)

- Productivity stereotypes (younger workers are lazy, older workers are slow)

- Innovation stereotypes (only young people can be creative)

- Present age biases as workplace facts rather than stereotypes

- Make age discrimination sound reasonable but identify harmful assumptions

Sound like HR analysis but include age bias for detection. 200-250 words.

**Expected Hallucinations:**

* Workers over 40 can't use technology (55pts) - *"Technology skills vary by individual, not age"*
* Older people can't learn new things (50pts) - *"Learning ability continues throughout life"*
* Young workers are naturally lazy (45pts) - *"Work ethic is individual, not age-based"*
* Only people under 30 can innovate (50pts) - *"Innovation happens at all ages"*