

Contents

MK1	2
MK2.....	4
MK3	5
MK4	6
MK5 Strict	8
MK5 Medium.....	10
MK5 Lenient.....	12

Johnathan, could you please explain the evolution of each of the prompts? What was wrong with previous prompts and how did it improve on the next version?

MK1

“I am building a structured database of battery-related academic offerings for students interested in careers in the battery industry. For each school I give you, extract only credit-bearing academic courses, academic certificates, or CEU programs that are explicitly or likely related to battery topics. Qualified matches include any program that teaches battery science, battery management systems, battery manufacturing, energy storage systems, lithium-ion batteries, electrochemical systems, solid-state batteries, redox flow batteries, thermal runaway mitigation, or closely associated technologies. Do not include programs focused on general electronics, basic circuit design, electric vehicles, automotive maintenance, or non-credit research unless they are integrated into a credit-bearing course. If you find a course that appears to be battery-relevant but does not explicitly confirm battery content, you must still include it in the output row. These may involve energy storage, electrochemical devices, or sustainability systems with possible battery components. You must scan the full school website including all subpages, course catalogs, CEU pages, and department listings. You are also required to use Open Syllabus and any PDF catalogs linked from the domain. Return the result as a single row of plain CSV data with no header row. The first column must contain the full name of the school. The second column must contain either Yes, No, or Website timeout. From the third column onward, list each detected course in its own separate column. Do not include any column labels, headings, headers, section titles, summaries, markdown, tables, or formatting. Return exactly one CSV row and nothing else.”

The MK1 prompt was first brainstormed when Aaron had a Google Colab code that searches the internet for whatever results desired by checking the top ten search links. We then realized we could utilize this code in order to check the validation of the learning language models (LLMs) that were web scraping for us. We then grabbed results we knew were confirmed correct and then grabbed results produced for us by the LLMs. The issue we encountered often with the LLMs was hallucinations or the creation of false data through its own creativity and its heuristic grouping of similar information. In order to determine if a course returned by Chat GPT was a hallucination or not, we fed the courses produced by the LLM and asked Aaron’s code to search for courses like them online. If results were returned back it meant that the course did exist, but if nothing was returned it meant that the course was made up by Chat GPT. Though in order to get to this point, I had to tweak Aaron’s code first to do what I needed it to do. The major issue was that I do not know how to code whatsoever. Instead I had to get creative and ask Chat GPT to essentially code for me through a process I’ve coined as “string to code”. I make sure I tell Chat GPT what I want and how I want it and then I upload the code that was previously used.

MK2

“I am building a structured database of battery-related academic offerings for students interested in careers in the battery industry. For each school I give you, output exactly one row of plain CSV with no headers, labels, bullet points, dashes, narratives, summaries, explanations, line breaks, or extra formatting. The first column must be the full name of the school. The second column must be either Yes, No, or Website timeout. From the third column onward, list each detected credit-bearing academic course, certificate, or CEU program as a separate column in the format COURSECODE: Course Title (for example: MECHENG 565: Battery Systems and Control). Output ONLY a single CSV row and nothing else. Do not add any preface, explanation, or content before or after the CSV row. If no qualifying programs are found, output the school name, then No. If the website cannot be loaded or times out, output the school name, then Website timeout. If a program or course does not have a code, use its official title only as a CSV column. Return only the CSV row and nothing else.”

It is not as simple as it seems to prompt engineer. Mentioned earlier was the statement that I simply uploaded the code into Chat GPT and that was the conclusion, but this is not accurate at all. MK1 produced many errors which was result of Chat GPT taking leniency with the prompt given to it. By leniency I mean the creative aspect of Chat GPT becomes too strong and starts to generate information not asked for or it takes liberties not requested. In order to by pass this I gave strict enforcement to not do anything unrequested. By strict enforcement I mean I used vocabulary that was extremely direct and left no room for error.

MK3

“You are forbidden from generating any output except for a single row of plain, valid CSV with no headers, quotes, bullet points, dashes, narrative text, line breaks, markdown, summaries, explanations, or extra formatting. The CSV row must begin with the full official name of the school, followed by a single word Yes, No, or Website timeout in the second column. Starting with the third column, list only the exact official names of all detected credit-bearing academic courses, certificates, or CEU programs that explicitly teach battery science, operation, diagnostics, or manufacturing. Only include offerings that directly focus on topics such as battery chemistry, battery materials, electrochemical energy storage, battery management systems (BMS), embedded battery control, SOC/SOH estimation, thermal runaway, lifecycle analysis, battery modeling, battery simulation, QA protocols, battery pack integration, anode or cathode design, redox processes, or solid-state or lithium-ion battery systems in a scientific or engineering context. Exclude any courses that only mention batteries in passing, general electronics, electrical maintenance, circuit design, power systems or energy storage courses not specifically about batteries, solar/battery hybrid content where batteries are secondary or not explicitly studied, and any course or program focused on electric vehicles (EVs), EV systems, propulsion, integration, diagnostics, infrastructure, hybrid electric vehicles, smart mobility, or automotive battery systems. If the course or program does not clearly and specifically teach battery-focused science or engineering, it should not be included. If no valid courses or programs are found, output only the school name and No. If the website times out, output only the school name and Website timeout. You must never output any narration, explanations, lists, quotes, or any additional lines before or after the CSV row. If you do not comply, your output will be discarded.”

Yet again MK2 failed and I moved onto MK3. The differences between each model type are virtually identical except for filters placed into the code and the changing of prompts which are fed into Chat GPT. MK3 I come to the conclusion that I do not understand code very well as others meanwhile Chat GPT knows nearly all internet bound knowledge with code included in the mix. I then told Chat GPT I didn't mind understanding what it's doing as long as it channels its effort into coding itself based on my English string prompt instead of explaining the code to me. Some ways this was accomplished was by telling Chat GPT to use vocabulary that it will process as strict. After this new implement MK3 worked better than one and two.

MK4

"You are forbidden from generating any output except for a single row of plain, valid CSV with no headers, quotes, bullet points, dashes, narrative text, line breaks, markdown, summaries, explanations, or extra formatting. The CSV row must begin with the full official name of the school, followed by a single word Yes, No, or Website timeout in the second column. Starting with the third column, list only the exact official names of all detected credit-bearing academic courses, certificates, or CEU programs that explicitly teach battery science, operation, diagnostics, or manufacturing. Only include offerings that directly focus on topics such as battery chemistry, battery materials, electrochemical energy storage, battery management systems (BMS), embedded battery control, SOC/SOH estimation, thermal runaway, lifecycle analysis, battery modeling, battery simulation, QA protocols, battery pack integration, anode or cathode design, redox processes, or solid-state or lithium-ion battery systems in a scientific or engineering context. Exclude any courses that only mention batteries in passing, general electronics, electrical maintenance, circuit design, power systems or energy storage courses not specifically about batteries, solar/battery hybrid content where batteries are secondary or not explicitly studied, and any course or program focused on electric vehicles (EVs), EV systems, propulsion, integration, diagnostics, infrastructure, hybrid electric vehicles, smart mobility, or automotive battery systems. If the course or program does not clearly and specifically teach battery-focused science or engineering, it should not be included. If no valid courses or programs are found, output only the school name and No. If the website times out, output only the school name and Website timeout. You must never output any narration, explanations, lists, quotes, or any additional lines before or after the CSV row. If you do not comply, your output will be discarded."

MK3 did not fulfil all requests and I was left unsatisfied. Other than strictness I realized there were other ways that Chat GPT understood itself better than I did. I realized that since it likes to take liberties I would have it ask permission before taking those liberties. The way this is done is by asking Chat GPT if it understands everything I tell it to do and then I ask Chat GPT to ask me clarifying questions in order to better its understanding further. Chat GPT then asks me questions that it would have most likely have just carried out if I had not asked it. This helps filter any errors through leniency, but the last issue remains. Strictness was not yet at its full potential, the vocabulary being used was not enough to stop Chat GPT from taking liberties. Instead I asked it for a technical reason why it was behaving the way it was in order to better understand it. Then I ask it for any technical work arounds based on the issue it perfectly just described to me. For some reason this crosses some kind of logic

issue with the LLM and allows it to fix itself better than when a human does it. If performed by a human the LLM will simply keep reproducing the same mistake, but if done by itself the issue will then more likely be fixed.

MK5 Strict

"DO not generate any output except for a single row of plain, valid CSV with no headers, bullet points, dashes, narrative text, line breaks, markdown, summaries, explanations, or extra formatting. "

"The CSV row must begin with the full official name of the school, followed by a single word Yes, No, or Website timeout in the second column. "

"Starting with the third column, list only the exact official names of all detected credit-bearing academic courses that have an official university course code (such as 'MECHENG 565: Battery Systems and Control'). "

"Do not include topics, modules, specializations, certificates, programs, or any item that does not have a university course code. "

"Only include offerings that directly focus on battery chemistry, battery materials, electrochemical energy storage, battery management systems (BMS), embedded battery control, SOC/SOH estimation, thermal runaway, lifecycle analysis, battery modeling, battery simulation, QA protocols, battery pack integration, anode or cathode design, redox processes, or solid-state or lithium-ion battery systems in a scientific or engineering context. "

"Exclude any courses that only mention batteries in passing, general electronics, electrical maintenance, circuit design, power systems or energy storage courses not specifically about batteries, solar/battery hybrid content where batteries are secondary or not explicitly studied, and any course or program focused on electric vehicles (EVs), EV systems, propulsion, integration, diagnostics, infrastructure, hybrid electric vehicles, smart mobility, or automotive battery systems. "

"If the course or program does not clearly and specifically teach battery-focused science or engineering and does not include an official university course code, it should not be included. "

"If no valid courses are found, output only the school name and No. If the website times out, output only the school name and Website timeout. "

"Do not output any narration, explanations, lists, quotes, or any additional lines before or after the CSV row."

MK4 did well, but after using the methods previously described I developed MK5 Strict. The only difference from the previous model was that it was given a list of keywords that were considered “battery related/specific” by our definition and was told to enforce the keywords and nothing else. It worked well, but perhaps too well. Suddenly we weren’t being returned as many courses as before.

MK5 Medium

"DO NOT generate any output except for a single row of plain, valid CSV with no headers, bullet points, dashes, narrative text, line breaks, markdown, summaries, explanations, or extra formatting. "

"The CSV row must begin with the full official name of the school, followed by a single word Yes, No, or Website timeout in the second column. "

"Starting with the third column, list only the exact official names of all detected credit-bearing academic courses that have an official university course code (such as 'MECHENG 565: Battery Systems and Control'). "

"Identify any credit-bearing course, certificate, or CEU program that provides meaningful instruction related to batteries. "

"Focus on offerings that include: battery management systems (BMS); lithium-ion or solid-state batteries; battery degradation, diagnostics, or modeling; battery safety systems, charge/discharge cycles, or thermal control; circuits or systems that directly integrate battery technology; power electronics involving battery storage; solar + battery system installations or lab-based battery work; UPS systems or rechargeable device architecture with embedded battery use. "

"You may also include: courses that emphasize portable battery devices, battery packs, or engineering contexts where batteries are actively used or controlled. "

"Do not include: general electronics, circuits, or energy systems that only mention batteries in passing; courses with minor battery references not central to the curriculum; any content primarily focused on: electric vehicles (EVs), EV propulsion, infrastructure, or diagnostics, smart mobility, hybrid electric vehicles, or automotive systems. "

"The result must be a course or program that explicitly engages with batteries in a way that teaches engineering, diagnostics, integration, control, or modeling—not simply using them as components. "

"If no valid courses are found, output only the school name and No. If the website times out, output only the school name and Website timeout. "

"Do not output any narration, explanations, lists, quotes, or any additional lines before or after the CSV row."

Nothing was necessary to change other than the prompt yet again. We did the same way as we transformed strict to medium leniency.

MK5 Lenient

"DO NOT generate any output except for a single row of plain, valid CSV with no headers, bullet points, dashes, narrative text, line breaks, markdown, summaries, explanations, or extra formatting. "

"The CSV row must begin with the full official name of the school, followed by a single word Yes, No, or Website timeout in the second column. "

"Starting with the third column, list only the exact official names of all detected credit-bearing academic courses that have an official university course code (such as 'MECHENG 565: Battery Systems and Control'). "

"Include any course, certificate, or CEU program that is related to batteries—whether directly or indirectly. "

"Include anything that references or involves: battery management systems (BMS); lithium-ion or solid-state batteries; battery degradation, modeling, chemistry, or QA; portable energy storage; rechargeable device design; battery-powered systems; electrical systems with battery input; UPS systems; battery use in labs or projects; general electronics or energy systems that involve battery use; solar + battery installations or circuits using batteries. "

"Do not include any course that is primarily about renewable energy, natural resources, sustainability, or environmental science, unless batteries are the core subject. "

"Exclude all courses focused on electric vehicles (EVs), automotive battery systems, EV charging, e-mobility, smart mobility, automotive engineering, or transportation. "

"Look for credit-bearing academic courses, certificates, or CEU trainings only. Skip non-credit workshops or EV-focused programs. "

"If no valid courses are found, output only the school name and No. If the website times out, output only the school name and Website timeout. "

"Do not output any narration, explanations, lists, quotes, or any additional lines before or after the CSV row."