

# KOREAN NOVEL GLOSSARY MAKER

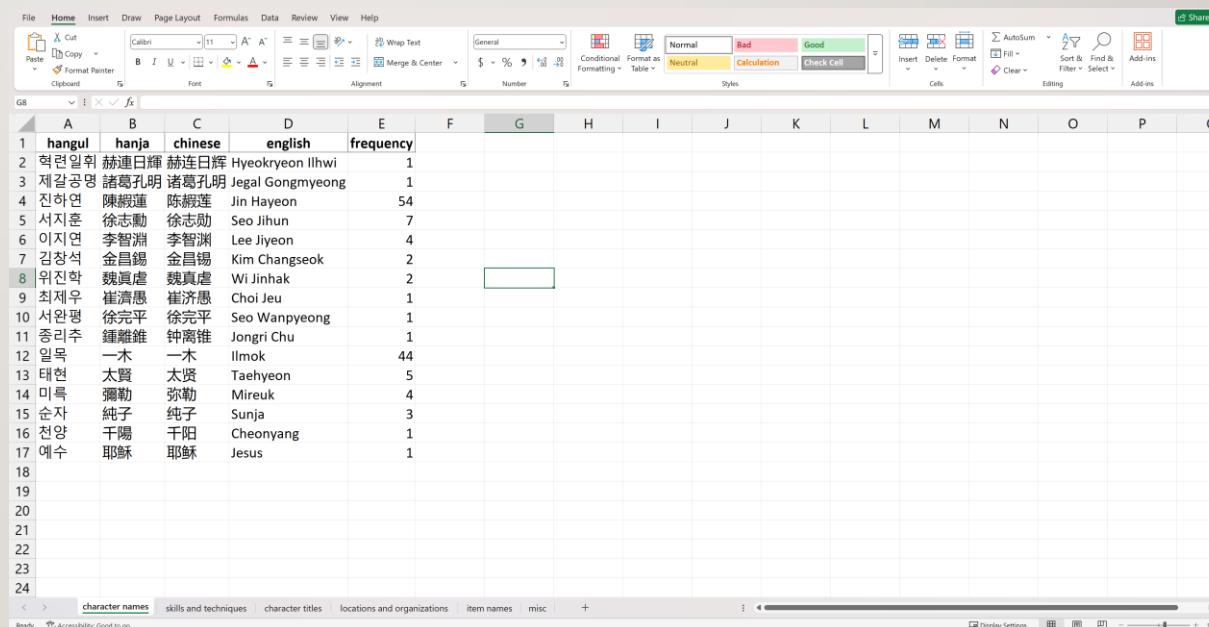
---

BY: FOODIEMONSTER007

# WHAT DOES GLOSSARY MAKER DO?

---

- This tool helps you create a translation glossary spreadsheet for Korean novels, to improve the quality of your machine translations. Here is an example of the output:



The screenshot shows an Excel spreadsheet titled "character names" with the following data:

	hangul	hanja	chinese	english	frequency
2	혁련일휘	赫連日輝	赫连日辉	Hyeokryeon Ilhwii	1
3	제갈공명	諸葛孔明	诸葛孔明	Jegal Gongmyeong	1
4	진하연	陳毅蓮	陈毅莲	Jin Hayeon	54
5	서지훈	徐志勳	徐志勋	Seo Jihun	7
6	이지연	李智淵	李智渊	Lee Jiyeon	4
7	김창석	金昌錫	金昌锡	Kim Changseok	2
8	위진학	魏真虐	魏真虐	Wi Jinhak	2
9	최제우	崔濟愚	崔济愚	Choi Jeu	1
10	서완평	徐完平	徐完平	Seo Wanpyeong	1
11	종리주	鍾離錐	钟离锥	Jongri Chu	1
12	일목	一木	一木	Ilmok	44
13	태현	太賢	太贤	Taehyeon	5
14	미륵	彌勒	弥勒	Mireuk	4
15	순자	純子	纯子	Sunja	3
16	천양	干陽	干阳	Cheonyang	1
17	예수	耶穌	耶稣	Jesus	1
18					
19					
20					
21					
22					
23					
24					

# REQUIREMENTS

---

- 1) **Python:** This is a programming language that makes the tool work. You'll need to install it from the official website (<https://www.python.org/downloads/>). Make sure to check the 2 boxes that say "Add Python to PATH" and "pip" during installation. This step lets your computer easily find and run the program.
- 2) **Google API Key:** This is a special password that connects the program to Google Gemini AI. You can get a free one from the Google Cloud Console (<https://console.cloud.google.com/>). Here are the instructions on how to do it: <https://www.youtube.com/watch?v=brCkpzAD0gc>
- 3) **Run "setup.py"** by double-clicking it before you proceed to ensure everything is properly set up. You only ever need to run it ONCE. You should see the message "SETUP SUCCESSFUL!" in the command prompt window that pops up.

# INSTRUCTIONS

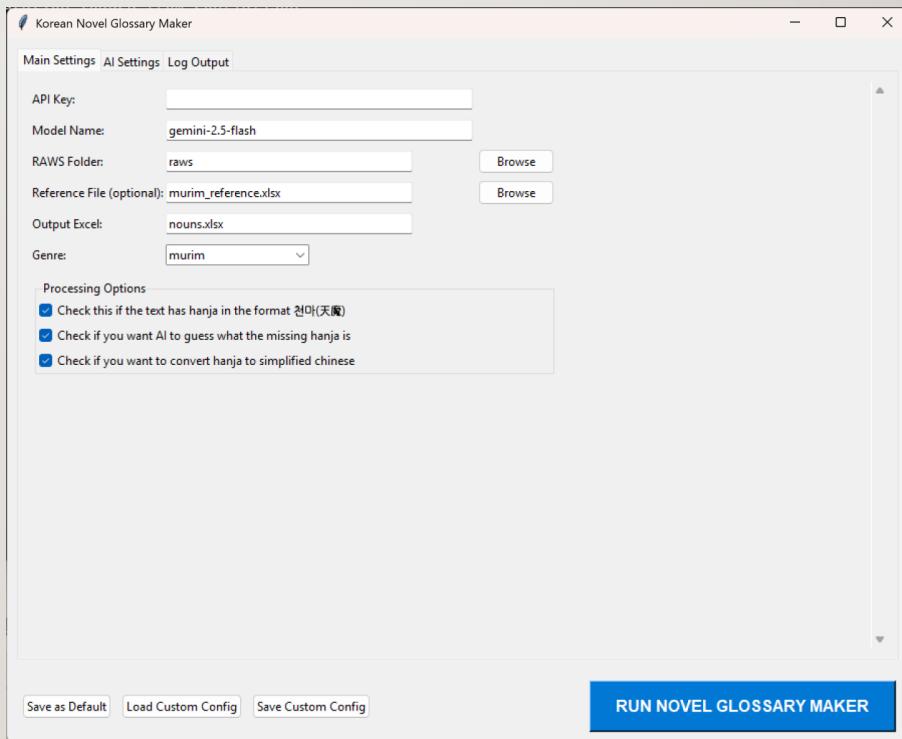
---

- Put your novel chapter Korean raws in .txt file format inside a folder and copy it into the main directory. I have provided a folder named "raws" as an example.
- You can name the folder anything you want, but it must not have spaces. E.g. “Absolute\_Regression” or “Solo-Leveling” are fine but not “Return of the Blossoming Blade”

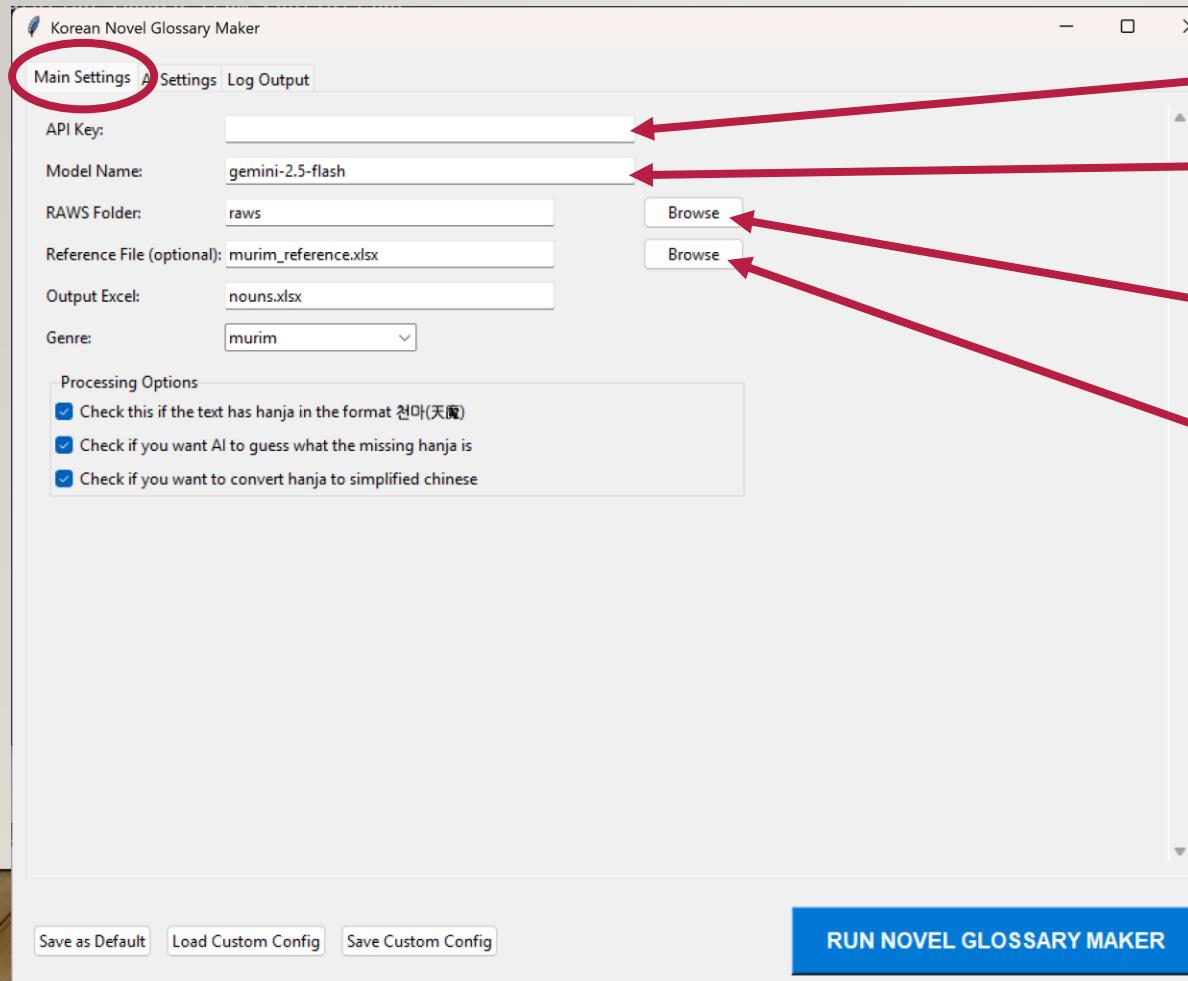
# INSTRUCTIONS

---

- Start the Program: Find the program file called **GLOSSARY MAKER.py** and double click it. You should see the following screen:



# INSTRUCTIONS – MAIN SETTINGS TAB



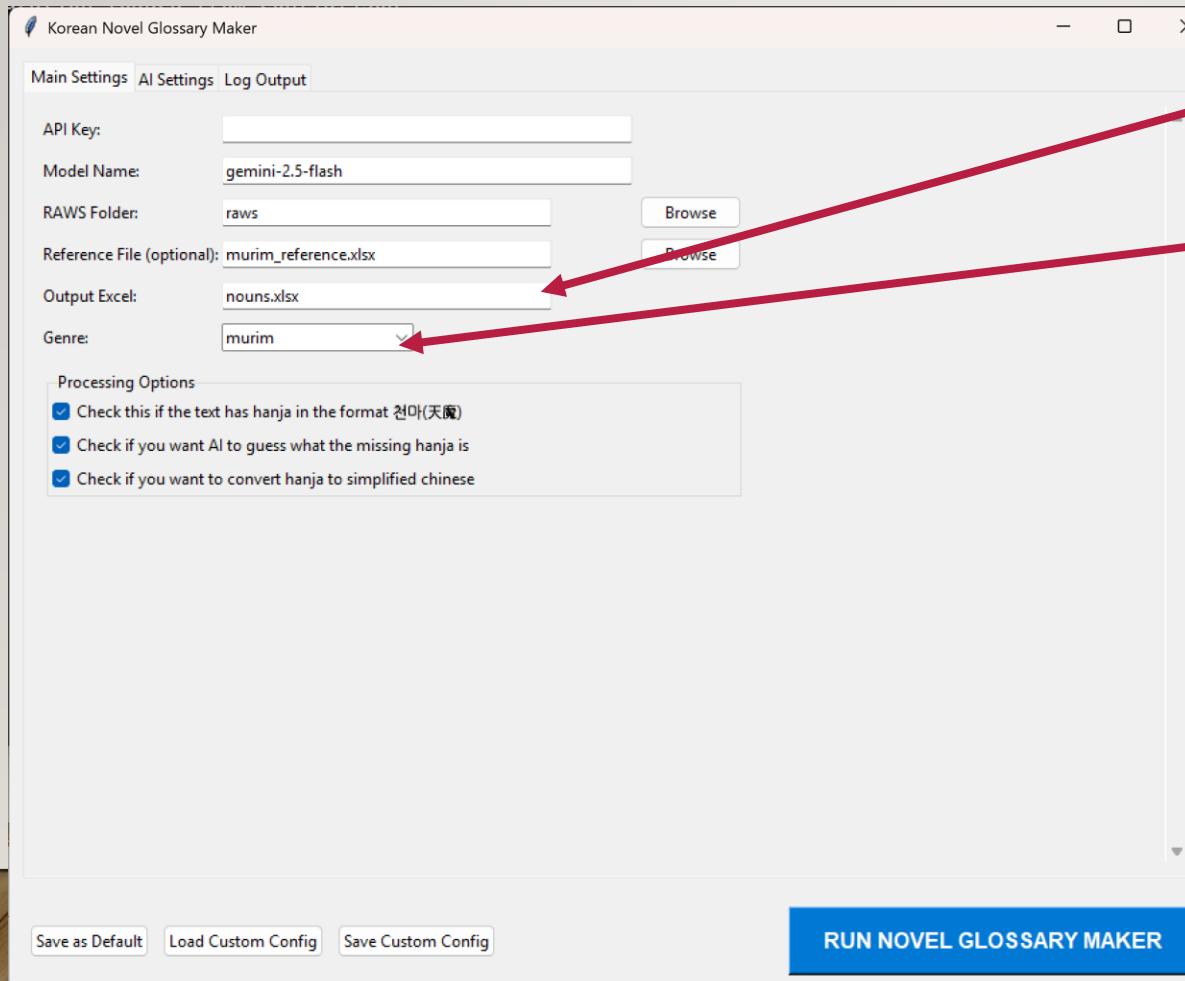
Paste the Google API key you got earlier here

Choose your favorite Gemini AI model.  
Note not all models are available for free usage.

Type the folder name or browse to the folder where  
you saved your Korean novel .txt files

If you are making a glossary for a murim novel, leave  
this as is. If you have made a glossary manually  
before, you can use your own reference excel  
spreadsheet, but note that it must follow the same  
format as murim\_reference.xlsx. If the novel does  
not belong to the murim genre and you don't have  
your own glossary yet, delete the text and leave this  
blank

# INSTRUCTIONS – MAIN SETTINGS TAB



Type your desired glossary file name here. Make sure it ends with .xlsx

Select the genre of your novel. We have:

**murim**: Chinese martial arts novel

**rofan**: romance fantasy novel

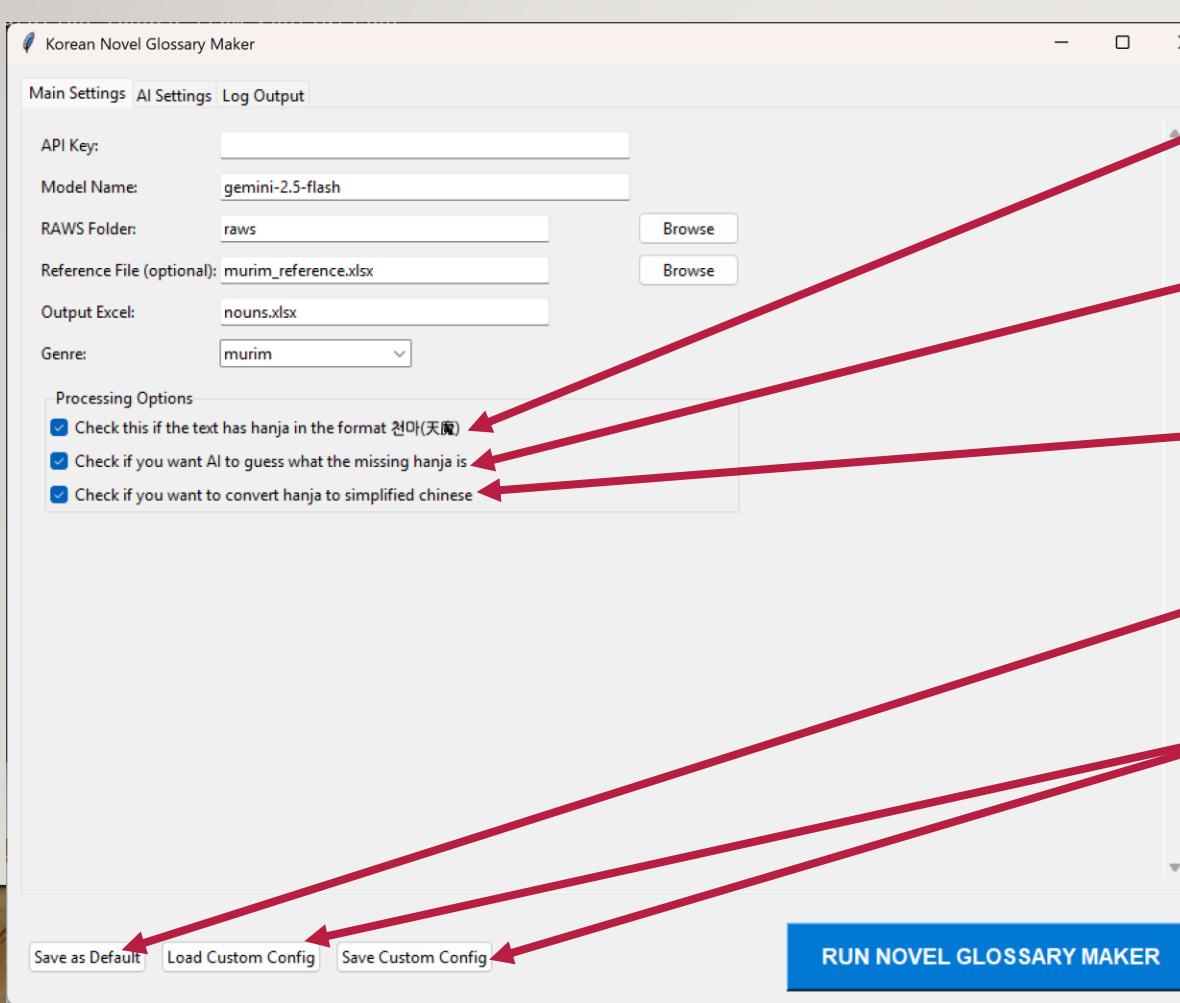
**modern**: modern day novel

**game**: MMO/VRMMO novel or system novel

**westfan**: western fantasy novel

**dungeon**: modern day novel with dungeons or towers

# INSTRUCTIONS – MAIN SETTINGS TAB



Check this only if your novel has hanja in brackets next to hangul

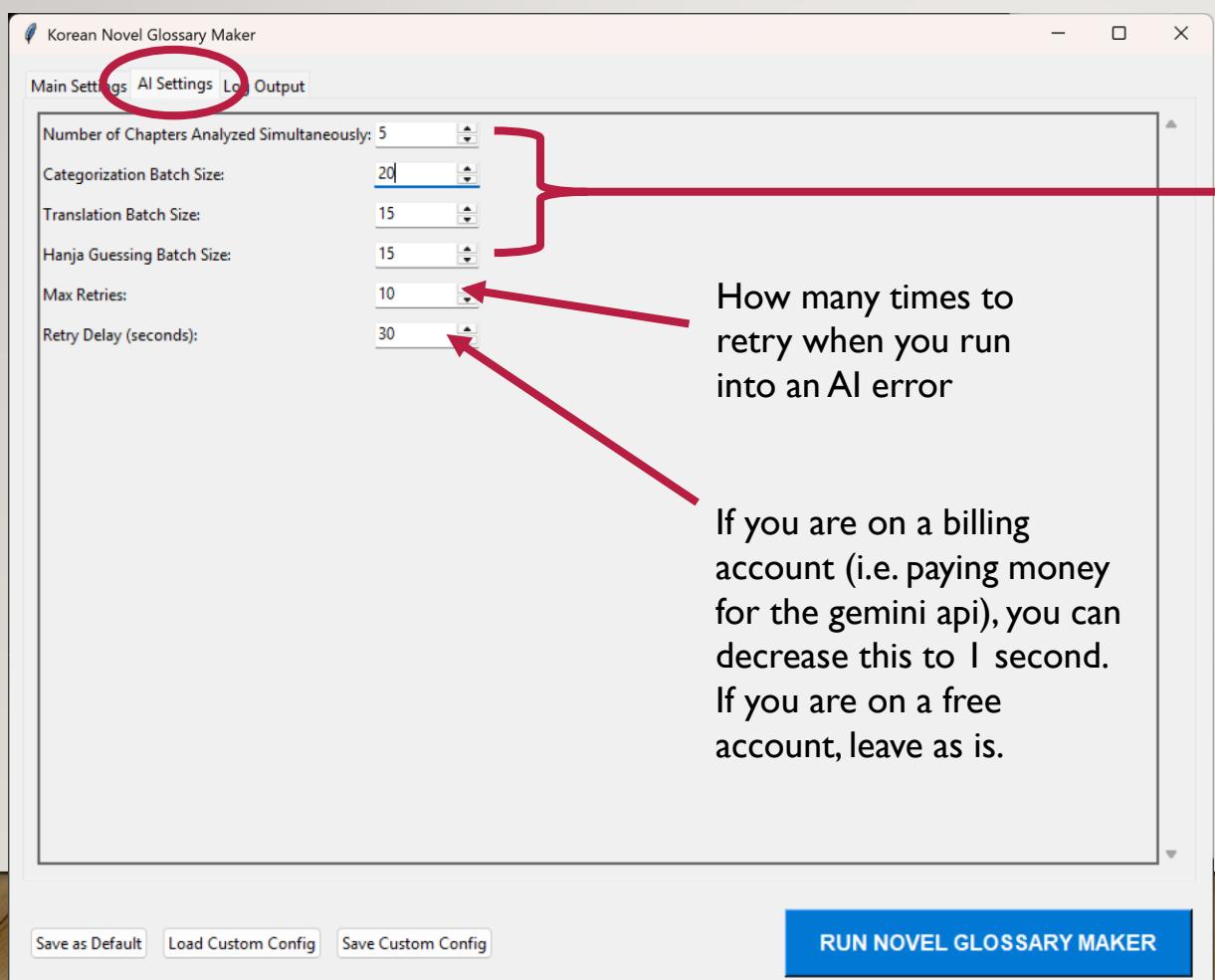
Not all glossary terms identified will have hanja. Check this if you want AI to guess the hanja

Check this if your novel has hanja, you asked AI to guess the hanja, AND you can read Chinese

Click this to save your current settings as default

Click these to save your current settings as a new custom config file, or load a custom config file that you saved previously

# INSTRUCTIONS – AI SETTINGS TAB



Increase these numbers if you wish to use fewer AI credits and save money

Decrease these numbers if you keep seeing the MAX\_TOKEN error in the output, or your output file looks strangely truncated

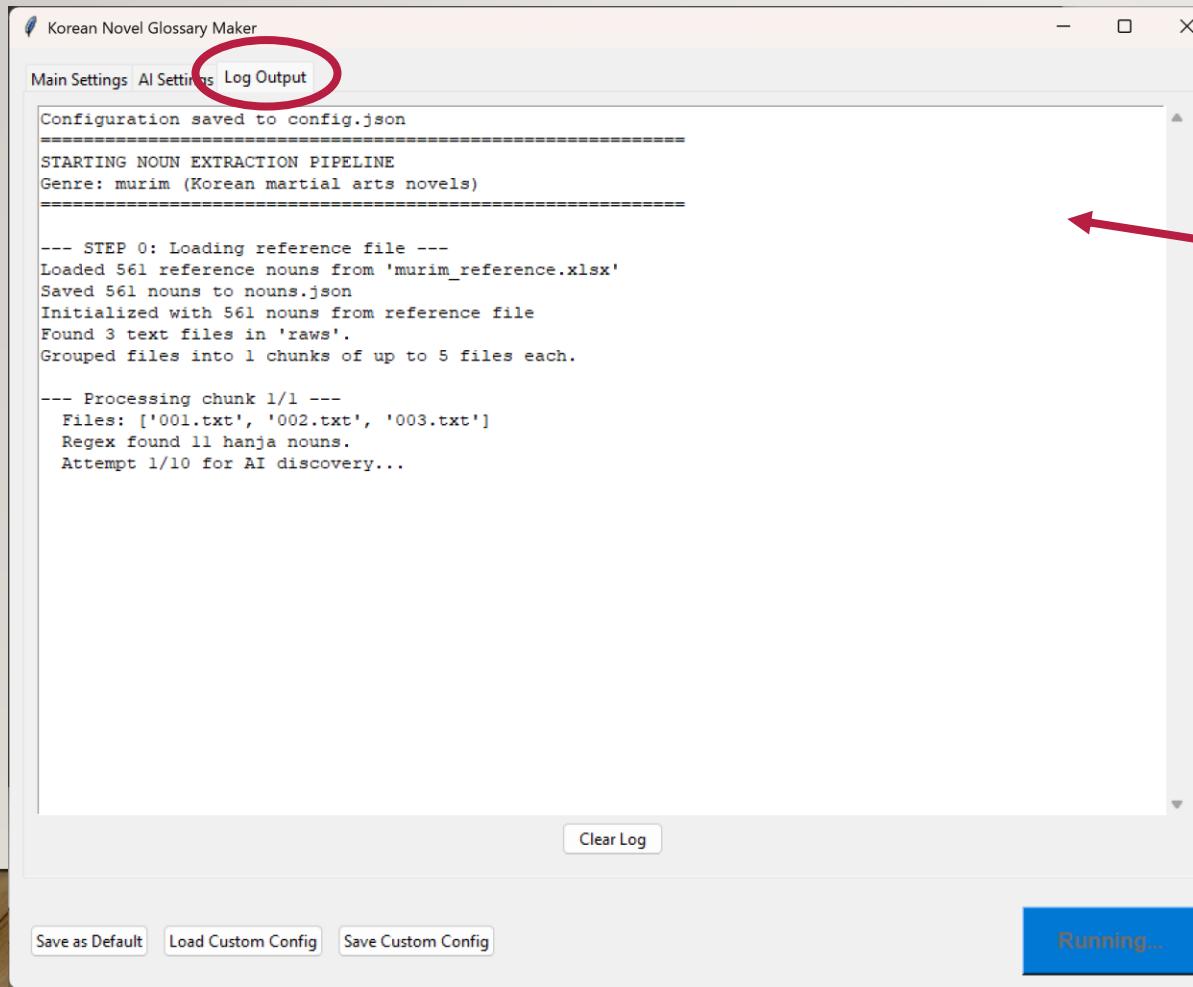
**Number of Chapters Analyzed Simultaneously:** How many novel chapters are analyzed at the same time. Decrease this if the novel chapters are long.

**Categorization Batch Size:** How many nouns are put into categories (names, locations, items) at the same time

**Translation Batch Size:** How many nouns are translated at the same time

**Hanja Guessing Batch Size:** How many nouns are submitted for hanja guessing at the same time (if enabled)

# INSTRUCTIONS – LOG OUTPUT TAB



Observe your glossary  
making progress

Watch out for errors

# PIPELINE COMPLETED SUCCESSFULLY!

---

- You should see a file named **nouns.json** in the folder. If you wish to continue adding nouns to the same novel (e.g. you made a glossary for chapters 1-10, and now you're updating it for chapters 11-20), this file will be necessary.
- nouns.json will be **overwritten** whenever you run the code, so if you are working on multiple novels, remember to **BACKUP** this file.
- You should also see **two new excel spreadsheets**. One of them has the terms in different categories in separate worksheets for easier visualization, while the other has every term on **ONE** worksheet for easier MTL-ing.
- You probably want to **check** the generated glossary spreadsheets manually and change any mis-translated terms and names. AI is stupid, it happens.

# HAPPY MTL-ING!

---

CREDITS: FOODIEMONSTER007