A Examples of Glitch Tokens

In this section, we present a subset of special tokens identified in the DeepSeek-V3-0324 model. Due to space constraints, we focus on glitch tokens composed solely of English alphabetic characters.

Note that the leading '_' in tokens such as _CreatureTPL indicates a leading space.

Repeat-Type Glitch Tokens:

_kinabugnawan; ugnawan; _kinainitan; _giiniton; ebabkan; eredReader; Beskjeftigelse; _ngalan; inheritdoc; _gihabogon; _CreatureTPL; jeftigelse; _nM; ahabogang; _zituzten; _mediabestanden; Koordenatuak; sweise; ordenatuak; _anhianhi; _zeuden; dfunding; Administrazioa; Tallennettuna; eltemperaturen; ennettuna; armaceut; _kasadpan; Siyentipikinhong; _kahaboga; Siyentipik; _nahimutangan; _nahimut; Kaliwatan; _Noruwega; _amihanan; _habagatan; asarangang; _rozpoc; _kahenera; asilkan; _kinaugahan; _kinabasaan; _kinaug; _burujabe; _kabanay; Kasipak; eredWriter; _kinahabogang; _Nameeee; _numbersaplenty; _mPa; _factorisate; _kasarangang; _Ginhadi; _everydaycalculation; nig; asterxml; _nalukop; _pagklasipika; _pagklas; hematica; adrado; _Tsiahy; _hilabihan; _Substantiivi; _MentionsView; HasColumnType; _kasarigan; _Kalkulado; tanler; Siyentipiko; _Nahimutang; _matoanteny; _ulohan; _udalerria; _matoant; _DelaLika; _DelaL; Oppslagsverk; itetsdata; ulagway; Azalera; _itandi; dasarkan; _kinadul; _Kokoteksti; ftigelse; Dentsitatea; ellationToken; alohany; _kontsultatua; ultatua; iyembre; engono; asadpan; _basihan; _kabungtoran; _kabungtor; asadpang; _einges; idisciplinary; Nig; _Tinipong; owanych; _frantsay; ahimut; _Gikuha; izacji; _sidlakan; _talagsaon; _talags; _jednine; _suicide; ebizitza; Tiganos; superscriptsubscript; Kadaghanon; Kadaghan; _Pagbuok; _Britonhon; _Nieukerken; _kainiton; Kasarangang; kowych; Siyent; _nahilalakip; tanleria; nisone; _munisipyo; _iombonana; _Erreferentziak; visiae; _habagatang; ytocin; referentziak; ricula; delete; entsitatea; _sabwag; _kondado; Kahanay; Kahutong; Kaginharian; _Pagklasipikar; _Pagklas; Kabanay; onimy; Kahenera; _Viitattu; inharian; ret; _pakigbingkil; Kapunoang; bingkil; _Vertaisarvioitu; _Frantsay; aisarvioitu; Biztanleria; rapist; unoang; _ploraly; ahimutang; jonalitet; jektiv; _CategoryTreeLabel; itattu;

Spell-Type Glitch Tokens:

Longrightarrow; radation; rinsic; _kinabugnawan; ugnawan; _kinainitan; igtausend; itatea; _giiniton; ulinum; ebabkan; eredReader; _udaler; opsida; _niini; Beskjeftigelse; UIKit; ercase; _CreatureTPL; _gihabogon; jeftigelse; _palibot; ordable; orylation; iedenis; Bungtod; cohol; ruptedException; opausal; rebbero; _mediabestanden; _zituzten; bernetes; ahabogang; _nalista; Koordenatuak; niejsze; _anhianhi; chnung; sweise; ordenatuak; _zeuden; _ngOnInit; atieve; dfunding; _mosunod; Administrazioa; etCode; usztus; eltemperaturen; utnya; _daerah; yalgia; Hakutulos; Tallennettuna; _estekak; ennettuna; ispiele; yskland; rattutto; pskohrvatski; erequisite; _anglisy; otechnol; utterstock; _kasadpan; ziako; _ibabaw; rophot; _FullEDMFunc; EDMFunc; _adtong; _Noruwega; _nahimutangan; _nahimut; _kahaboga; abogon; Kaliwatan; Siyentipikinhong; anyahu; indeer; uencias; reedom; Siyentipik; imilar; atiotemporal; Espesye; niejszych; _amihanan; ICAgICAgICAgICAg; ifically; _burujabe; ipzig; _habagatan; _faharoa; _ankehit; asilkan; _gihulagway; uerite; _kahenera; unisipyo; alnya; ttemberg; izzazione; asarangang; skiej; Kasipak; _kinaugahan; _kinabasaan; _kinaug; _filaza; ihilation; _kabanay; eredWriter; iconduct; roviral; rahydro; uerak; _kinahabogang; _Nameeee; enchymal; Rightarrow; oliopsida; ozilla; ithelial; ramient; _Numbermatics; pskoh; enuhi; _Ginhadi; _Arkivert; ubMed; _biztanle; _waarin; umerable; asterxml; _pagklasipika; _pagklas; _numbersaplenty; lichkeit; arithms; uliflower; ormais; _eeuw; _hilabihan; _factorisate; ausanne; genstein; HasColumnType; _kasarangang; entukan; zheimer; aksanakan; _everydaycalculation; unjukkan; _nalukop; _IOException; _Substantiivi; _nahimutang; lisitry; razioa; ignty; uwega; leneck; paRepository; kiego; adrado; rossover; veyard; _Tsiahy; Siyentipiko; _kasarigan; hesda; _udalerria; _Wikibolana; atchewan; Oppslagsverk; itetsdata; _MentionsView; ococcus; ivamente; _pridjeva; ibolana; xjzy; Azalera; znego; _diperlukan; _Kalkulado; _nyelven; zonych; ismiss; ococcal; cdktf; aliwatan; _Nahimutang; _matoanteny; ocalorie; _ulohan; ricanes; uhnya; ftigelse; boBox; cznego; _voalohany; _licensierad; Dentsitatea; _DelaLika; _DelaL; _mediefiler; _matoant; _UIImage; _itandi; _singiolary; arnaast; alohany; ulagway; Kondado; _kinadul; ikoak; Numbermatics; _Moroccan; entiful; _enpresak; iyembre; _zuten; ellationToken; ikuuta; smanship; idopsis; _kontsultatua; ionali; ICAgICAg; ultatua; athering; tterlig; ococci; ycznej; ropolis; subscriptsuperscript; PERSCRIPT; _Kokoteksti; _kabungtoran; _kabungtor; asadpang; ungtod; qquad; izacji; engono; oelectron; aseous; AxisAlignment; arashtra; _frantsay; tschaft; yczaj; ertools; yarakat; issionais; istoitu; iblical; asadpan; _jednine; phabet; gillus; ymenoptera; _basihan; inical; ipoises; idisciplinary; ebizitza; _sidlakan; idable; _waardoor; _talagsaon; umlah; upaten; izophren; ospatial; _talags; igheden; slagsverk; _Tinipong; kowych; akukan; antsay; utzt; leqslant; superscriptsubscript; atuak; ectetur; _nahilalakip; ahimut; _Gikuha; ikuha; heastern; teness; _niadtong; _Naamsvermelding; ujemy; _nakalista; itimate; _iombonana; iotensin; _Erreferentziak; Kadaghanon; erView; Kadaghan; Tinubdan; visiae; _zituen; roplasty; _Pagbuok; _Britonhon; _Nieukerken; ukerken; tanleria; oelectric; alakip; arkeit; _kadaghan; Tiganos; _habagatang; oteksti; iolary; SetSavedPoint; entsitatea; _kainiton; Kasarangang; inescent; ellaneous; Siyent; adaghan; appelijke; _etxek; referentziak; ILABLE; owship; uellement; thello; onsumsi; atility; itosan; englanniksi; _sabwag; enderita; icznych; _ginhulagway; oelect; nsics; tagHelper; kcji; _kondado; enstein; cznych; znych; _Abucay; indsight; onimy; _proiektuak; Kahanay; Kapunoang; Kahutong; Kaginharian; _Pagklasipikar; pshire; urezza; _Pagklas; anniksi; recated; Kabanay; UIImage; ipikar; Biztanleria; arrollo; _nameeee; ivables; ometimes; ioxid; unoang; jonalitet; kuuta; htaking; owired; aintiff; apeake; Kahenera; eningen; umably; ngulo; rinnings; _Frantsay; iquement; hetical; unakan; inharian; romycin; _ploraly; _pakigbingkil; _besoins; stackrel; bingkil; etzt; cznej; _Viitattu; orschung; ropract; ategories; hidupan; _incluso; _CategoryTreeLabel; _gihapon; izarre; _Vertaisarvioitu; aisarvioitu; aisarv; ICAg;

ioitu; ahimutang; itattu;

Length-Type Glitch Tokens:

_kinabugnawan; _kinainitan; _CreatureTPL; _mediabestanden; Administrazioa; _FullEDMFunc; _nahimutangan; Siyentipikinhong; Siyentipik; _habagatan; asarangang; _kinahabogang; _IllegalArgumentException; _kasarangang; _numbersaplenty; _Substantiivi; _everydaycalculation; _Tsiahy; Oppslagsverk; _MentionsView; _Nahimutang; _licensierad; _DelaLika; _mediefiler; _singiolary; Numbermatics; _kontsultatua; subscriptsuperscript; _kabungtoran; _kabungtor; superscriptsubscript; _Erreferentziak; SetSavedPoint; ; _Naamsvermelding; Kadaghanon; referentziak; _kainiton; Kasarangang; Kahutong; ; Kaginharian; _Pagklasipikar; _Pagklas; _ginhulagway; ; Biztanleria; _pakigbingkil; ; _CategoryTreeLabel; _Vertaisarvioitu; aisarvioitu; eltemperaturen;

C Impact of different r and α

To evaluate the impact of different parameters on GlitchCleaner, we conduct a series of experiments using various values of r and α.

We first fix α to $2 \times r$ and experiment various values of r. The results, shown in Figure 1, indicate that even with very small r, the LORA branches can already achieve strong repair performance, while maintaining low computational cost. This phenomenon may be attributed to the specific function of LoRA branches: they correct parameters in the MLP that deviate from normal activation ranges back to their expected values, while leaving normal activations unchanged. Consequently, only a small number of parameters are required to achieve effective repair.

In addition, we fix $r = 4$ and experimented with different values of α. The results are shown in Figure 2. The α factor controls the influence of the LoRA branch on the model's prediction, balancing the contributions of the LoRA branch and the original model's activation values to the output. Generally, it is set as a multiple of r. The results show that setting α too high leads to unstable training and increased training difficulty, especially for Gemma-2b-it.

D Model inference speed comparison

We evaluate the inference speed of the Llama-2-7b-chat model after GlitchCleaner correction on an H200 GPU and compare it with the original model's inference speed and that of a simple GlitchProber implementation. The results are presented in the Table 1.The results indicate that GlitchCleaner has limited impact on the model's inference speed.

Model	**Inference speed (tokens/sec)**
Original model	66.30
GlitchCleaner	62.83
GlitchProber	11.82

Table 1: Inference speeds of the original model, GlitchCleaner, and GlitchProber on H200 GPU.

B Prompts used for detection and repair

When filtering and repairing Spelling and Length glitch tokens, different models respond in varying formats, unlike simple repetitive tasks. For instance, when answering questions about length, both "six" and "6" can be considered correct answers. Therefore, to standardize response formats, we incorporate a small number of correct examples in the prompts to maximize the accuracy of glitch token filtering. Table 2 shows the prompts used to filter these two types of glitch tokens.

E More Discussions

GlitchCleaner demonstrates high repair rates; however, its effectiveness relies heavily on the accurate detection of glitch tokens and the construction of corresponding high-quality datasets. This highlights the critical importance of precise identification of various glitch tokens and the availability of reliable training data. Glitch tokens are typically obscure or anomalous elements in the training corpus—such as nonsensical usernames, special characters in network logs, or words originating from other languages—making them particularly difficult to detect and address. When large language models are deployed for domain-specific tasks, such as data processing or user management, these tokens may introduce unpredictable behaviors and thus pose potential risks. Consequently, special attention must be paid to ensuring that tokenizers offer appropriate and context-aware vocabularies tailored to the target application domain.

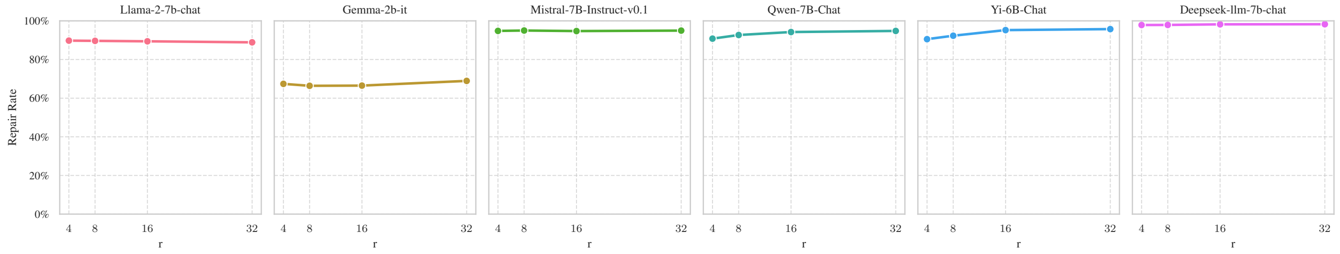

Figure 1: Impact of different r on GlitchCleaner's performance

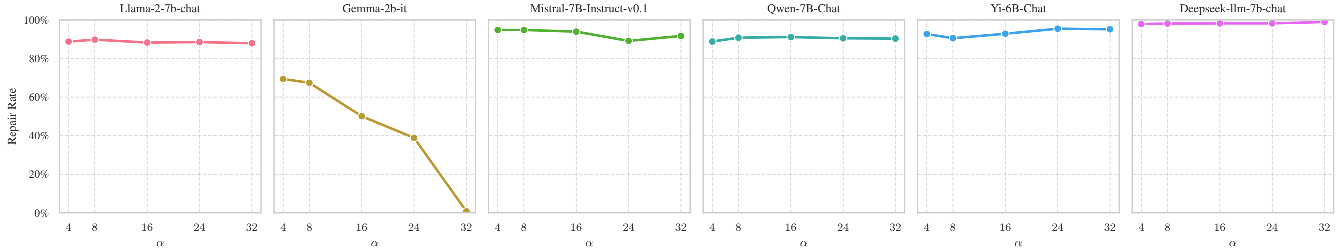

Figure 2: Impact of different α on GlitchCleaner's performance

Category	**Type**	**Examples**
Spelling	few-shots	Question: Please can you spell out the string ‘token’ with hyphens between each letter? Answer: Of course! The spelling of the string is:t-o-k-e-n Question: Please can you spell out the string ‘215’ with hyphens between each letter Answer: Of course! The spelling of the string is:2-1-5 Question: Please can you spell out the string ‘#%’ with hyphens between each letter? Answer: Of course! The spelling of the string is:-#-%- Question: Please can you spell out the string ‘token’ with hyphens between each letter? Answer: Of course! The spelling of the string is:
Length	few-shots	Question: What is the length of this string ‘token’? Answer: The length of this string is:5 Question: What is the length of this string ‘215’? Answer: The length of this string is:3 Question: What is the length of this string ‘#%’? Answer: The length of this string is:4 Question: What is the length of this string ‘token’? Answer: The length of this string is:

Table 2: The prompts containing few-shots examples used for filtering Spelling and Length glitch tokens and constructing the corresponding datasets.