

1. Fine-tuning LLM for summarization

1.1 Dataset and Task Selection

- **Dataset Name:** Yle News 2014-2020 ([Selkosuomi Parallel Corpus](#)).
- **Content:** The dataset consists of 8,261 pairs of standard Finnish news articles and their corresponding expert-written simplified summaries ("Selkosuomi").
- **Task:** The objective is text summarization and simplification. The goal is to fine-tune a model to condense complex news articles into easy-to-read Finnish summaries.

1.2 Fine-Tuning Technique Family

- **Technique:** [QLoRA](#) (Quantized Low-Rank Adaptation).
- **Implementation:** I used the Unsloth library to load the model in 4-bit quantization and attached LoRA adapters to specific layers (q, k, v, o, gate, up, down).

1.3 Personal Variation

- **Dataset-Specific Objective:** Instead of generic summarization, the model was trained to respond to a specific Finnish instruction: "*Tiivistä seuraava uutinen selkosuomeksi*" (Summarize the following news into plain Finnish).
- **Constraints:**
 1. **Hardware Limit:** The training had to fit strictly within 8GB VRAM (NVIDIA RTX 3070 Ti Laptop GPU).
 2. **Parameter Limit:** Max 3B model.
 3. **Training Time:** I wanted the training to take MAX 4 hours, but preferably less.

```
==((=====))==
Unsloth 2026.1.4: Fast Llama patching. Transformers: 4.57.6.
\\  /|  NVIDIA GeForce RTX 3070 Ti Laptop GPU. Num GPUs = 1. Max memory: 8.0 GB. Platform: Linux.
0^0/ \_/\ Torch: 2.10.0+cu128. CUDA: 8.6. CUDA Toolkit: 12.8. Triton: 3.6.0
\     / Bfloat16 = TRUE. FA [Xformers = 0.0.34. FA2 = False]
"-____-" Free license: http://github.com/unslotha/unsloth
Unsloth: Fast downloading is enabled - ignore downloading bars which are red colored!
```

Figure 1: System specifications.

2. Implementation & Results

Fine-Tuning Strategy and Justification

I chose [unsloth/Llama-3.2-3B-Instruct](#) because it is lightweight enough for local deployment on consumer hardware and it was already instruction tuned. I used Unsloth for the training pipeline because it optimizes memory usage and increases training speed by 2x compared to

standard Hugging Face implementations. Without Unsloth, I don't think this would have been possible, or at least I would have had to make drastic modifications.

Pipeline Description

1. **Preprocessing:** Raw .csv files were loaded and formatted into a chat template structure (user instruction + model answer).
2. **Configuration:** The model was loaded with 4-bit quantization. LoRA rank was set to 32 and alpha to 16 to balance performance with VRAM usage. "unsloth" gradient checkpointing reduced VRAM usage significantly.
3. **Training:** The model was trained on a subset of 800 examples for 1 epoch using UnslothTrainer.
4. **Inference:** The fine-tuned adapter was merged with the base model to generate summaries for new articles.

```
model = FastLanguageModel.get_peft_model(  
    model,  
    target_modules = ["q_proj", "k_proj", "v_proj", "o_proj",  
                      "gate_proj", "up_proj", "down_proj"],  
    lora_alpha = 16,  
    r = 32,  
    lora_dropout = 0,  
    bias = "none",  
    use_gradient_checkpointing = "unsloth", # saves lots of vram!!!  
    random_state = 3407,  
    use_rslora = True,  
    loftq_config = None,  
    max_seq_length = MAX_SEQ_LENGTH  
)
```

Figure 2: LoRA config.

Training and Evaluation Evidence

- **Input Data:** The model successfully ingested 800 training examples and 15 test examples. Only 15 for the validation, because I didn't want training to take too long. But something like a 800 / 200 split would obviously been better.

```

--- SAMPLE INPUT FOR TRAINING ---
<bos><start_of_turn>user
Tivistä seuraava uutinen selkosuomeksi:

Isopanda synnytti kaksoset Chengdun pandakeskuksessa maanantaina. Kaksoset olivat ensimmäiset tänä vuonna syntyneet. Kelin-nimininen isopanda synnytti maanantaina tyttökaksoiset Chengdun pandankasvatuskeskuksessa Kiinan lounaisosassa. Poikaset olivat maailman ensimmäiset tänä vuonna syntyneet pandakaksoiset. Poikasten kerrotaan olevan hyvässä kunnossa. Kaksosista vanhempi painaa 118 grammaa ja nuorempi 70 grammaa. Poikaset syövät vain millilitran maitoa kerrallaan ruokaillessaan, kertoo hoitajia haastatellut Kiinan valtiollinen CCTV. Pandakeskuksen emopanda tuli raskaaksi keinoisemminkään. Isopanda on luokiteltu erittäin uhanalaiseksi lajiksi. WWF:n mukaan luonnossa elää vain 1 800 isopandaa.<end_of_turn>
<start_of_turn>model
Kinin on syntynyt vuoden ensimmäiset pandakaksoiset. Pennut syntivät maanantaina Chengdun kaupungin keskukseissa, jossa kasvatetaan pandakarhuja. Luonnossa elää alle 2 000 isopandaa, joita se on vaarassa hävitä maailmasta kokonaan.
<end_of_turn>
```

Figure 3: Sample input for training

- **Memory Management:** The memory footprint was successfully contained, although my GPU:s VRAM was pushed to its limits (7.7 / 8 GB in use almost constantly).
- **Loss Curve:** The training loss demonstrated a downward trend, indicating the model learned the summarization format. At around step 100/150 overfitting began, but I found the overfitted model to be better. As that is what we actually want for this type of task.

```

... The model is already on multiple devices. Skipping the move to device specified in `args`.
==((=====))= Unsloth - 2x faster free finetuning | Num GPUs used = 1
  \\ /| Num examples = 800 | Num Epochs = 3 | Total steps = 150
  0^0/ \_/\ Batch size per device = 1 | Gradient accumulation steps = 16
  \     / Data Parallel GPUs = 1 | Total batch size (1 x 16 x 1) = 16
  "-__-" Trainable parameters = 24,313,856 of 3,237,066,752 (0.75% trained)
Unsloth: Will smartly offload gradients to save VRAM!
[150/150 1:14:21, Epoch 3/3]

Step Training Loss Validation Loss
30      1.943100    1.902443
60      1.754700    1.846373
90      1.636800    1.824506
120     1.482700    1.827940
150     1.468500    1.826902
```

Figure 4: Training and evaluation.

One Concrete End-to-End Example

- **Input:** A complex [news article](#) about a startup called "Donut Lab" and their new battery technology.

- **Model Output:** The finetuned model gets all the important facts, and the language is indeed simple “selkosuomi”. The base models summaries are also a lot longer. Example below.
- **Result:** The model successfully identified key facts (100,000 cycles, 400 Wh/kg, and also the working temperature range, which the base model didn't get) and synthesized them into valid, simple Finnish.

Inference Comparison Between the Base Model and Finetuned Model

- **Base Model:**

Suomalaisista startup-yrityksistä Donut Labin uusi akkuteknologia voisi muuttaa radikaalisti kilpailuasetelmia maailman akkumarkkinoilla. Yrityksen mukaan kyse on maailman ensimmäisestä teolliseen tuotantoon valmiista kiinteän olomuodon akusta.

Donut Labin akkuteknologia on herättänyt epäilyjä asiantuntijapiireissä, mutta yrityksen mukaan sen akkuja asennetaan jo virolaisen Verge Motorcyclesin valmistamiin sähkömoottoripyöriin.

Verge-moottoripyörän valmistaja on kuitenkin ajautunut talouskriisiin, ja yrityksen mukaan heidän akkunsa energiatehys on 400 Wh/kg, se latautuu täyteen alle viidessä minuutissa ja se kestää jopa 100 000 lataussykliä.

Sähkötekniikan professori Pertti Kauranen kertoo videolla, että näillä ominaisuuksilla suomalaisakku ohittaisi kirkkaasti nykyiset litiumakut.

Kiinasta tyrmäys väitteille on kommentoitu myös. Maailman suurimpiin kuuluvan kiinalaisen akkupalvelustajan Svoltin toimitusjohtaja Yang Hongxin tyrmäsi on herättänyt epäilyjä. Hongxinin mukaan kyse on huijauksesta.

Aalto-yliopiston akkukemian ja -materiaalien professori Tanja Kallio epäilee väitteiden paikkaansapitäävyyttä.

- **Finetuned Model:**

Suomalainen startup-yritys Donut Lab on esittänyt uuden akkuteknologian. Yritys kertoii, että sen akku on kiinteän olomuodon akku. Kyse on maailman ensimmäisestä teollisesta tuotantoon valmiasta kiinteän olomuodon akusta. Donut Labin akku on 400 Wh/kg, se latautuu täyteen alle viidessä minuutissa. Akku kestää jopa 100 000 lataussykliä. Donut Labin akku toimii lähes täydellä teholla 30 asteen pakkasessa ja yli sadan asteen kuumuudessa. Akku on myös paloturvallinen ja ekologinen. Donut Labin akku on parempi kuin nykyiset akut. Donut Labin akku on 400 Wh/kg, nykyiset akut ovat 150 Wh/kg.

Discussion of Limitations

- **Hardware Bottleneck:** The strict 8GB VRAM limit forced the use of a smaller dataset subset (800 rows) and aggressive gradient accumulation, potentially limiting the model's total knowledge retention.
- **Training Time:** Training took 1 hour 14 minutes, with such a small training dataset. Using the whole dataset (8261 document pairs), would have indubitably resulted in better results. However, then training time would have taken almost ten times longer with my debilitated GPU.
- **Model Size:** The 3B model turned out to have great performance already from the start. However, its summaries were too long and the language was not “selkosuomi”.
- **Evaluation:** Because I didn't want to train the model all day long I had to limit the validation set to a laughable 15. Therefore, the evaluation losses are not very accurate.

Suomalaistaustaisen startup-yrityksen Donut Labin uusi akkuteknologia voi muuttaa radikaalisti kilpailuasetelmia maailman akkumarkkinoilla, jos yrityksen hurjat väitteet uuden akun ominaisuuksista pitävät paikkaansa.

Nykyään Kiina on selvässä johtoasemassa akkuteknologian kehittämisessä ja akkujen valmistuksessa.

Lappeenrannan-Lahden teknillisen yliopiston sähkötekniikan professorin Pertti Kaurasen mukaan Donut Labin Yhdysvaltain CES-messuilla esittelemä uusi kiinteän olomuodon akkuteknologia voisi muuttaa tilannetta Euroopan ja Suomen hyväksi.

– Tällainen teknologia varmasti kelpaisi eurooppalaisille autonvalmistajille. Ne saisivat sen avulla kilpailuetua etenkin kiinalaisiin sähköautonvalmistajiin, professori Kauranen sanoo.

LUT-ylipiston sähkötekniikan professori Pertti Kauranen kuvattuna yliopiston sisätiloissa.

Aavaa kuvien katselu

Professori Pertti Kaurasen mielestä akkuteollisuuden voimatasopaino voi järkyä, jos Donut Labin akkuteknologia lyö läpi. Kuva: Kalle Schönberg / Yle
Myös Kiinan hallitseva asema nykyisen akkuteknologian kriittisten raaka-aineiden, kuten lithiumin, louhinnassa ja jalostamisessa voisi horjua, koska Donut Labin mukaan sen akun raaka-aineet ovat yleisesti saatavissa.

– Pääsisimme eroon geopoliittisista riippuvuuksista, jos raaka-aineita olisi saatavissa helposti mistä tahansa, Kauranen sanoo.

Hurjat ominaisuudet

Donut Lab esitti akkuteknologiaansa tammikuun alussa Yhdysvalloissa Las Vegasin CES-messuilla. Yrityksen mukaan kyse on maailman ensimmäisestä teolliseen tuotantoon valmiista kiinteän olomuodon akusta.

Donut Labin mukaan sen akkuja asennetaan jo virolaisen Verge Motorcyclesin valmistamiin sähkömoottoripyöriin, joista ensimmäiset lähtevät asiakkaille lähikuukausien aikana.

Ylen tietojen mukaan Verge-moottoripyörän valmistaja on kuitenkin ajautunut talouskriisiin.

Donut Labin mukaan heidän akkunsa energiatehys on 400 Wh/kg, se latautuu täyteen alle viidessä minuutissa ja se kestää jopa 100 000 lataussykliä.

Yhtiön mukaan akku toimisi lähes täydellä teholla vielä 30 asteen pakkasessa ja yli sadan asteen kuumuudessa. Lisäksi akut olisi edullista valmistaa ja raaka-aineet saataisiin läheltä. Akut olisivat myös paloturvallisia ja ekologisia.

Sähkötekniikan professori Pertti Kauranen kertoo videolla, että näillä ominaisuuksilla suomalaisakku ohittaisi kirkkaasti nykyiset lithiumakut.

LUT-ylipiston sähkötekniikan professori Pertti Kaurasen mukaan suomalaisakun väitettyt ominaisuudet ovat huomattavasti paremmat kuin parhaimpien nykyakkujen ominaisuudet. Video: Kalle Schönberg/Yle

Kiinasta tyrmäys väitteille

Donut Labin kovat väitteet akkuteknologiastaan ovat herättäneet epäilyjä asiantuntijapiireissä.

Asiaa on kommentoitu myös Kiinasta. Maailman suurimpiin kuuluvan kiinalaisen akkuvalmistajan Svoltin toimitusjohtaja Yang Hongxin tyrmäsi kovin sanoin Donut Labin väitteet kiinalaisten tiedotusvälineiden haastattelussa. Hongxinin mukaan kyse on huijauksesta.

Myös Aalto-yliopiston akkukemian ja -materiaalien professori Tanja Kallio epäilee väitteiden paikkaansapitäävyyttä.

– Se tuntuu ihan uskomattomalta. Se rikkoisi kaikkia niitä fysiikan ja kemian periaatteita, joita itse tunnen, Kallio sanoo.

Teknologia tuntematonta

Donut Lab ei ole ilmoittanut, minkälaisista teknologiateista se käyttää uudessa akussaan. Yhden vihjeen teknologian mahdollisesta alkuperästä antaa Donut Labin osakkuusyrityksen Nordic Nano Groupin antamat tiedot.

Imatralla tehdasta käynnistävä Nordic Nano Group on kertonut, että se pystyy valmistamaan akkukennoja nanomassasta, joka on kehitetty saksalaisen tutkimuksen pohjalta.

Yrityksen toimitilat lumisessa maisemassa aamuhämärässä Imatralla.

Aava kuvien katselu

Nordic Nano Group on käynnistämässä tehdasta Imatralla. Kuva: Tanja Hannus / Yle
Nordic Nano Groupin toimitusjohtaja Esa Parjanen kertoi Ylelle vuoden 2024 lokakuussa, että nanomassa voi korvata autojen ja kännyköiden akuissa tyypillisesti käytetyn litiumin.

– Nämä nanoakkukennot kestävät kymmeniä tuhansia latauskertoja ja niihin mahtuu enemmän energiota. Ne ovat myös paloturvallisempia, eivätkä voi räjähtää, Parjanen sanoi.

Yrityksen mukaan se alkaa tuottamaan samasta aineesta Imatran-tehtaassa aurinkopaneeleja ja aurinkoenegiaa kerääviä ohutkalvopinnoitteita jo ennen kesää. Varsinaisen massatuotanto alkaa syksyllä.

Parjanen mukaan tehtaassa ei tuotettaisi akkukennoja. Hän ei kerro, missä niiden tuotanto mahdollisesti alkaa.

--- SUMMARY — ORIGINAL

Suomalainen startup-yritys Donut Labin on esittänyt uuden akkuteknologian, joka voisi muuttaa akkumarkkinoilla tilanteen Euroopassa ja Suomessa. Yrityksen mukaan akkunsa energiatihleys on 400 Wh/kg, se latautuu täyteen alle viidessä minuutissa ja se kestää jopa 100 000 lataussykliä.

Donut Labin akkua on koottu tehtävässä Imatralla, jossa käynnistää Nordic Nano Group. Yrityksen toimitusjohtaja Esa Parjanen on kertonut, että nanomassa voi korvata autojen ja kännypöiden akuissa tyyppillisesti käytetyn litiumin.

Akkuteknologian väitteet ovat herättäneet epäilyjä asiantuntijapiireissä, mutta Donut Labin on varmistanut, että akkunsa ominaisuudet ovat huomattavasti parempia kuin parhaimpien nykyakkujen ominaisuudet.

Kiinasta puolesta, maailman suurimpiin kuuluvan kiinalaisen akkuvalmistajan Svoltin toimitusjohtaja Yang Hongxin on tyrmäsi Donut Labin väitteille. Hongxinin mukaan kyse on huijauksesta.

--- SUMMARY — FINETUNED

Suomalainen startup-yritys Donut Lab esitti uuden akun CES-messuilla Yhdysvalloissa. Yritys kertoi, että akku on kiinteän olomuodon akku. Se on teollisen tuotannon valmiina. Akkuun on mahdollista lataa akun kymmeniä kertoja. Akku kestää kymmeniä vuosia. Akku on myös paloturvallista. Yritys kertoi, että akku on edullista valmistaa. Akkuun on mahdollista saada raaka-aineet helposti mistä tahansa. Donut Lab sanoo, että sen akku on parempi kuin nykyiset akut.