RECEIPT CONTEXTUALIZER

Make receipts speak.

Our Target
Mission and Goals

➔ Use the data of receipts for insights

➔ Expenses for each category of grocery

➔ Target group are customers, not companies

➔ Search products which are similar, but not necessarily in the same category

➔ No static database which must be maintained. Use natural language processing

I know my expenses thanks to this app!

The Receipt c
confusing and hard t

- Brand names
- Abbreviations
- Not standardized
- Not machine readable
- Every Store differently

RISPENTOMATE
0,840 kg x   2,69 EU
GEHACKTES 2.0
  Rabatt 30%
BIO FETA 45%
FETA PDO LAKT-FR
FLADENBROT
KRAUTSALAT
KNOBLAUCHOLIVEN
GRIE. PFEFFERON.
  2 Stk x   1,09
CURLY WURLY
  3 Stk x   0,59
PUEBLO BLUE

SUMME

Geg. VISA

* * Kundenb

Datum:
Uhrzeit:
Beleg-Nr.
Trace-Nr.
                Bezahl
                DKB Visa
Nr.   ###
VU-Nr.
Terminal-ID
Pos-Info
AS-Zeit 09.02.
AS-Proc-Code = 00 055
Capt.-Ref. = 0000
APPROVED
Betrag EUR

Tel.: 0451

UID Nr.: DE

RAUECHERLACHS
JA! MARKENBUTTER
  2 Stk x   1,39
CHIPSF. SOUR CRE
NATURALS BALSAM.
CASHEWS PIKANT
NIMM2 LACHGUMMI
N2 LG HEROES
  2 Stk x   1,19
ALTOBELLO BIANCO
RAUCHZART 40%
COLA 1,5
  3 Stk x   0,89
PFAND 0,25 EURO
  3 Stk x   0,25
PAULANER SPEZI
  3 Stk x   0,59
PFAND 0,25 EURO
  3 Stk x   0,25
VIRG. MOJITO
PFAND 0,25 EURO
HAPPY MARACUJA
HAPPYDAY J-BEERE
RECYCL. TOILETT.
TOILETTENPAP. 3LG
TODAY KOSMETIKTU
  3 Stk x   0,75
BENEFUL

REWE

REWE Markt
Sabine Klitsch oHG
Barbaraweg 7
06773 Gräfenhainichen
Tel.: 034953-859060
UID Nr.: DE285931933

		EUR
VKORN SONNENBATZ		2,19 B
GENUSS SCHEIBEN		2,58 B
2 Stk x	1,29	
POMELO HONEY		4,58 B
2 Stk x	2,29	
BIO MARGARINE		3,58 B
2 Stk x	1,79	
DINKEL-CRUNCHY		2,59 B
AUFSTRICH PAP.ZU		4,05 B
3 Stk x	1,35	

SUMME	EUR	19,57
Geg. BAR	EUR	20,00
Rückgeld BAR	EUR	0,43

Steuer %	Netto	Steuer	Brutto
B= 7,0%	18,29	1,28	19,5
Gesamtbetrag	18,29	1,28	19,5

TSE-Signatur: NtRhO+7khiJrXyPFv6+dKx3ghricjq
                   DqGiTQYY3yu7rF3JKx61bSfcfem056
                   6yXBCt2LP31QkSOyfMxCsImSf1MnSg
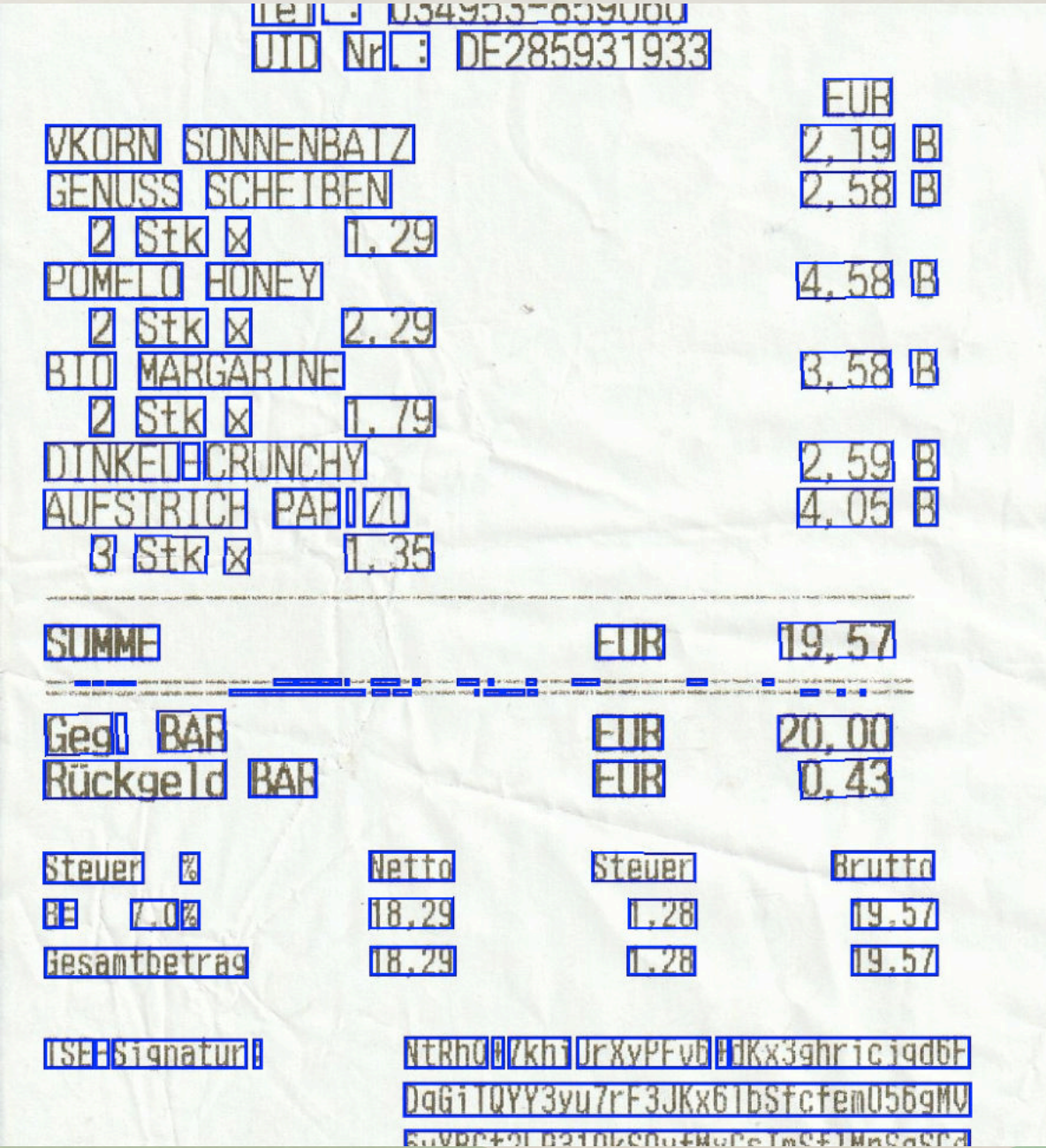   LeeV3tepW7YZdG031paFTYkI5uB2z
TSE-Signaturzähler: 1941613
TSE-Transaktion: 926203

Read Text from Image
Text recognition and segment parsing

Receipts:

Readout challenges:

- Heavily depends on image/receipt quality
- Non-uniform recognition over image
- Varying receipt structure per vendor

Use *Google Vision API* for 🌐
Optical **C**haracter **R**ecognition 👓

Match bounding boxes ☐←→☐
and clean data 🔄

ID	Product_abbr	Price
1	VKORN SONNENBATZ	2.19
2	GENUSS SCHEIBEN	2.58
3	POMELO HONEY	4.58
4	BIO MARGARINE	3.58
...	...	...

LLM as classifier

Embeddings for search

Performance evaluation

correct vs. meaningful
augmentation

$$Accuracy = \frac{Correct\ predictions}{All\ predictions}$$

ID	Product_abbr	Product_name	Main_category	Sub_category
1	BRESSO KR. PROV	Bresso Kräuter Provence	Obst & Gemüse	Frische Kräuter
2	VEG. SP. SCHNITTL	Vegetarische Schnitzel	Fleisch & Fisch	Fleischalternativen
3	ALTENBURGER FK	Altenburger Fettkäse	Käse, Eier & Molkerei	Käse & Käseersatz
4	REISPAPIER	Reispapier	Kochen & Backen	Backzutaten
…	…	…	…	…

(scanned: ~300 product abbreviations)

LLM accuracy:

Product name: **87%**

Main category: **84%**

Sub category: **82%**

Future work

Add feedback for LLM output

Generalize receipt processing (all vendors)

Include nutrition information

Update daily promotions

↓

Recommendations

Improve OCR:
- Preprocessing
- Preferably open-source
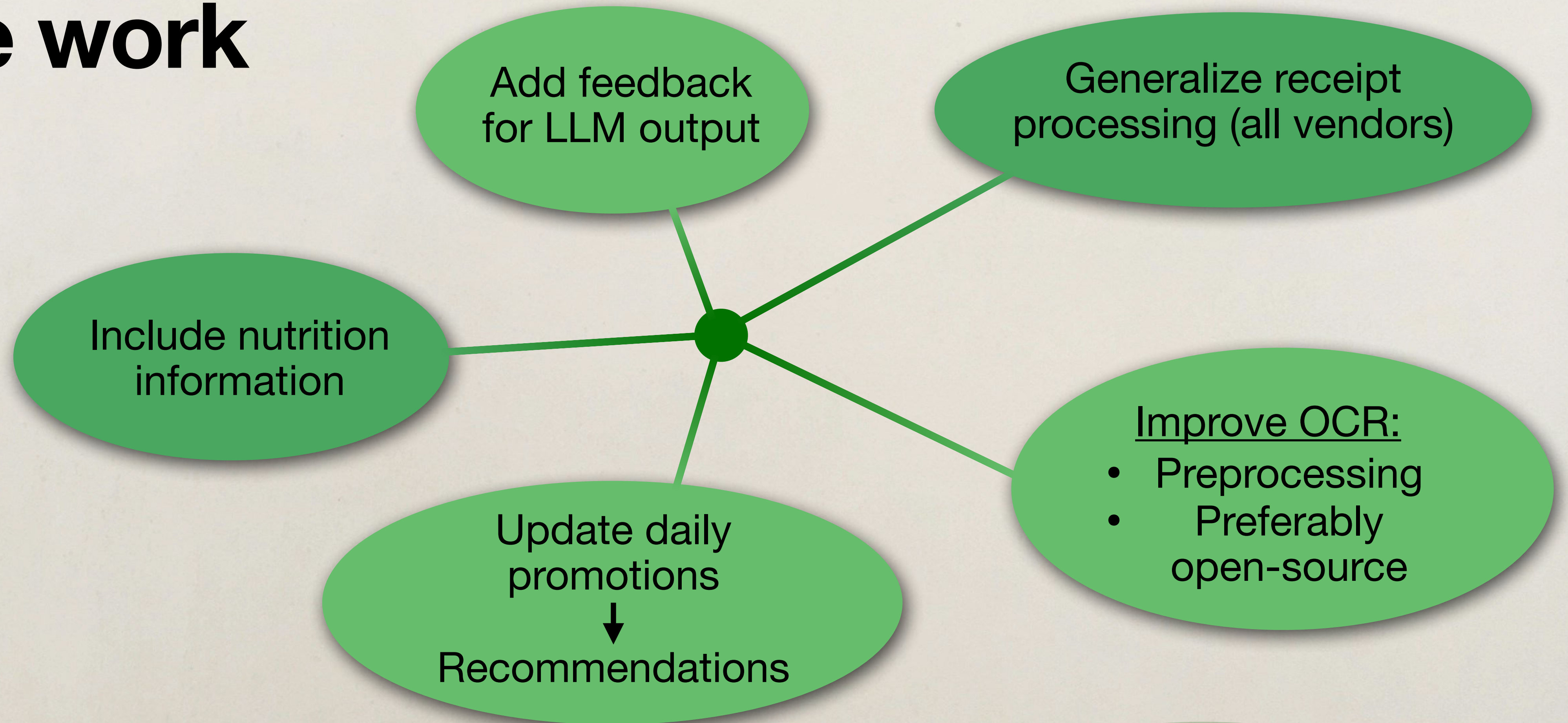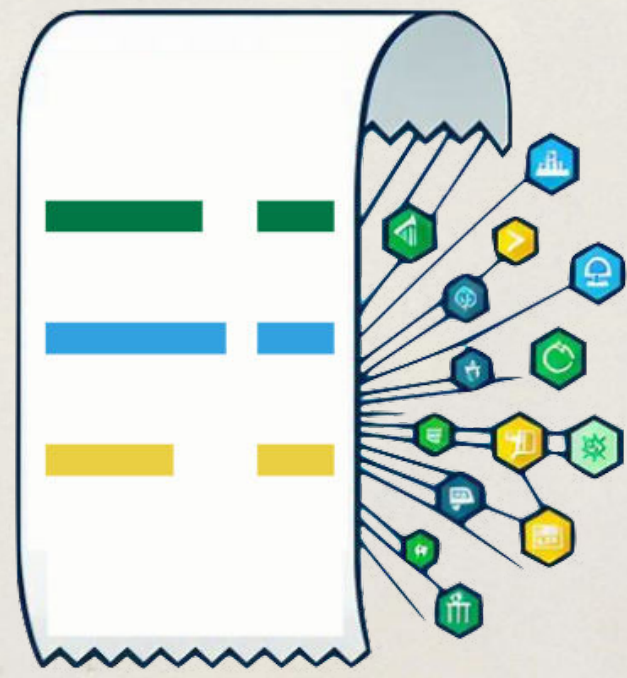
RECEIPT CONTEXTUALIZER

Thank you for your attention!

Tobias Pötzl, linkedin.com/in/tobias-pötzl

Leonard Kamps, linkedin.com/in/leo-kamps

Jakob Butlewski, linkedin.com/in/jakob-butlewski