X-Class

X-Ray Analysis for Tax Classification of Goods

X-Class

- Naïve Goal
  - Use X-ray images of containers to detect tax classification errors.
- Startup Challenges Program
  - Solving public problems and fostering startups
  - Two years project
    - going on
    - to be completed in January 2027
  - Inovia
    - https://inoviacorp.ai

Context

- An AI already detects tax code errors since 2014:
  - Sisam
  - Uses import declaration history
    - Tendencies of importers, suppliers, manufacturers, countries,
    - History of detected errors,
    - Description of goods,
    - **Not images**
  - Produces error probabilities item by item
  - Produce Probabilities of alternative tax codes
  - Justifies probabilities in natural language
- All bonded warehouses are forced to send scanner images to Brazil's Federal Revenue Secretariat through an API

Characteristics of the problem

- Difficulties
  - Overlap of goods
  - 10,000 codes
  - Millions of products
  - Very similar products
  - Labeling without location
  - No separation of items
- Extra obstacle
  - Sharing data with a startup

- Advantages
  - Constant Scale
  - Comprehensive labelling
  - Sisam as starting point

Real Goal and strategy

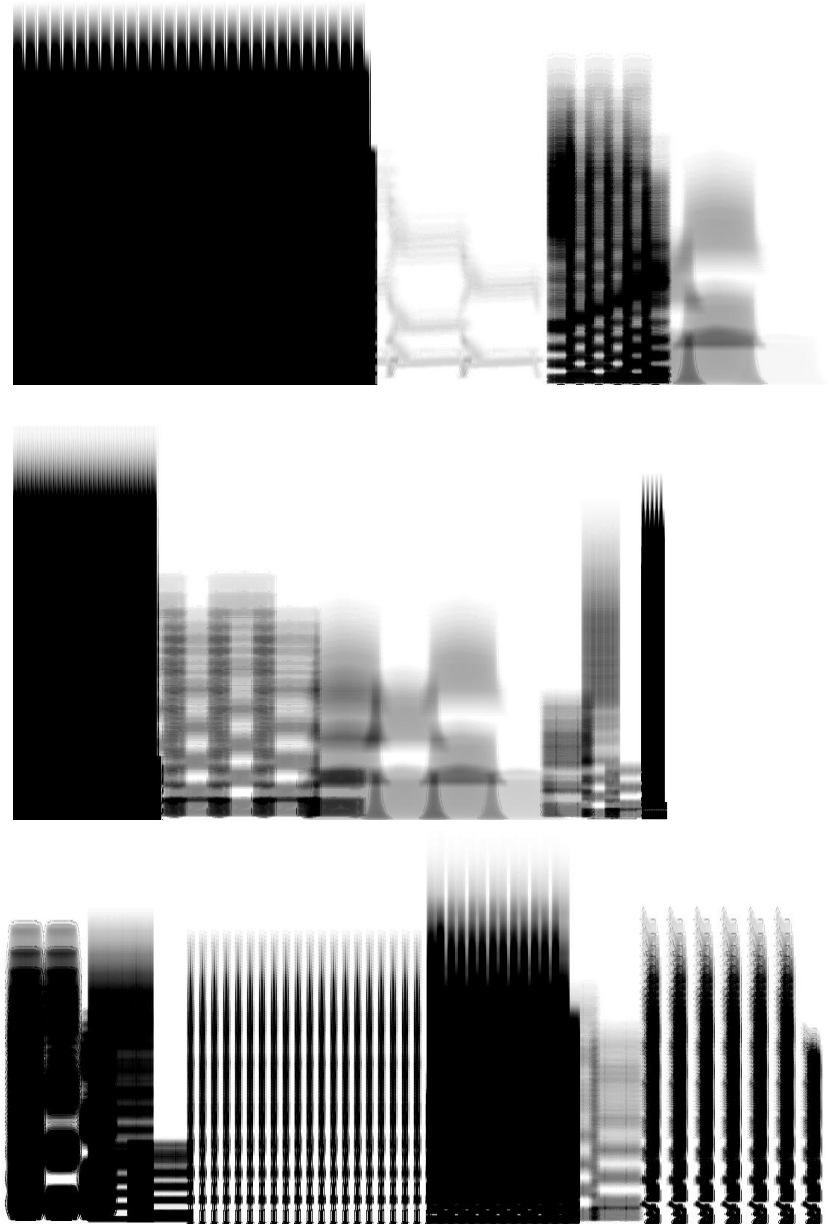
- **Real goals**
- Improve Sisam's probabilities using the images
- Provide images to be used inside Sisam's explanations

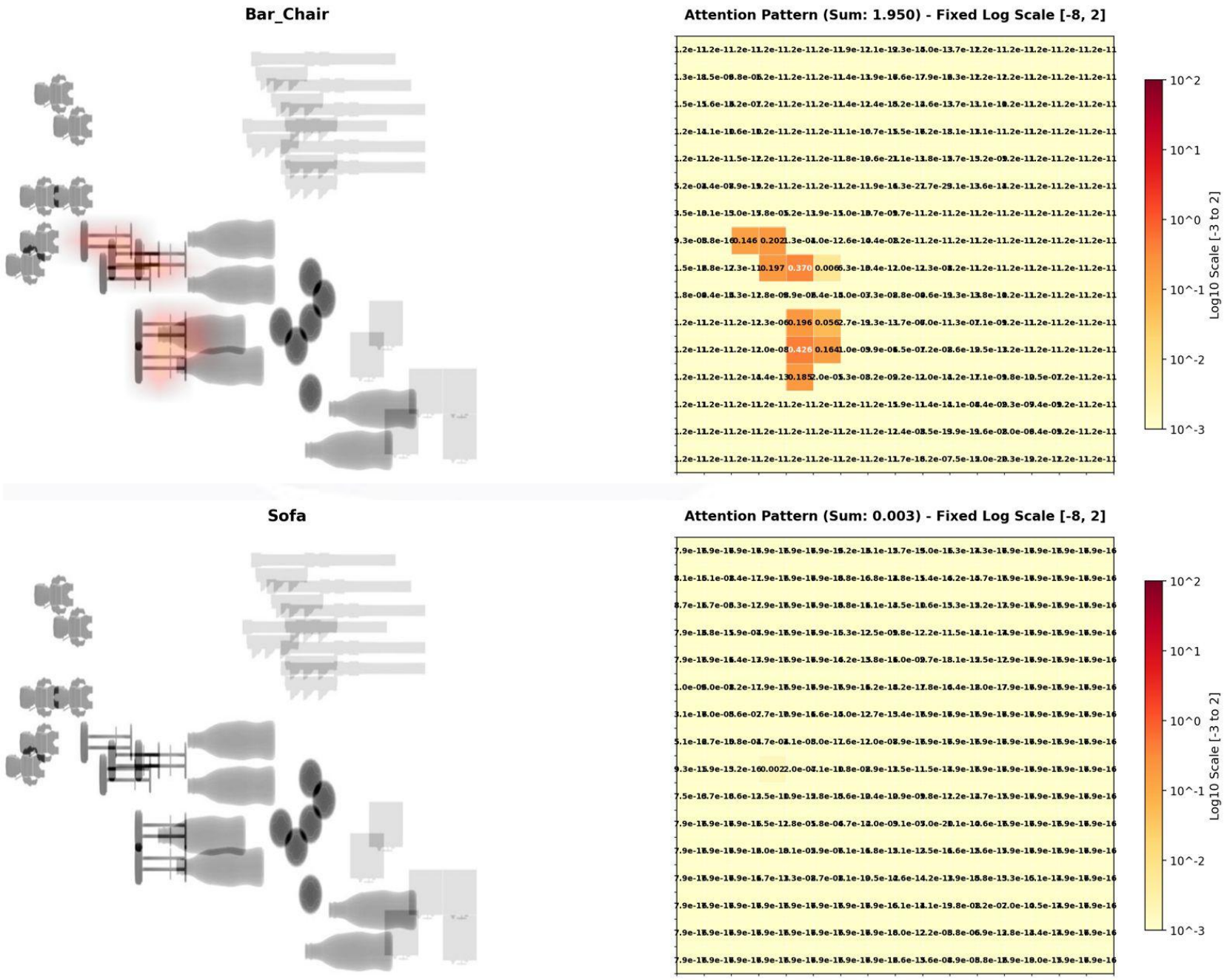
- **Strategy**
- Sisam reduces the number of candidates from 10000 to 10 with probabilities.
- X-Class applies Negotiated Weakly Supervised Object Localization
- X-Class produces probability distributions for the total weight for each tax code.
- Neural network
  - Resembles a transformer, isn't one.
- The probabilities of the two systems are combined
  - Monte Carlo Markov Chain Simulation
- Use artificial images

Synthetic data

- We cannot easily share the images
  We positioned shapes in a container
- We projected

- We can control complexity.
- Real data will be tested internally

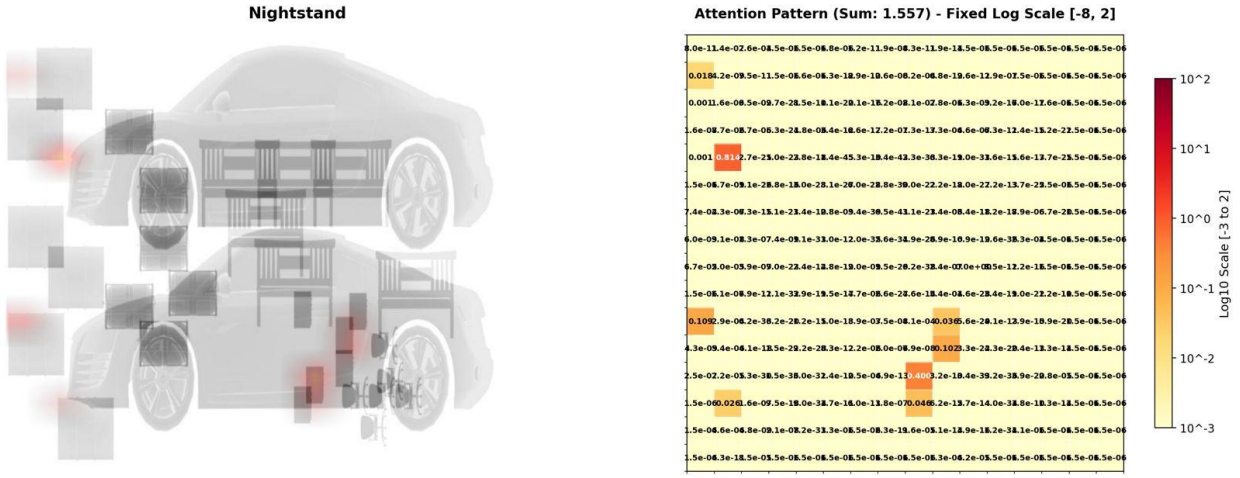


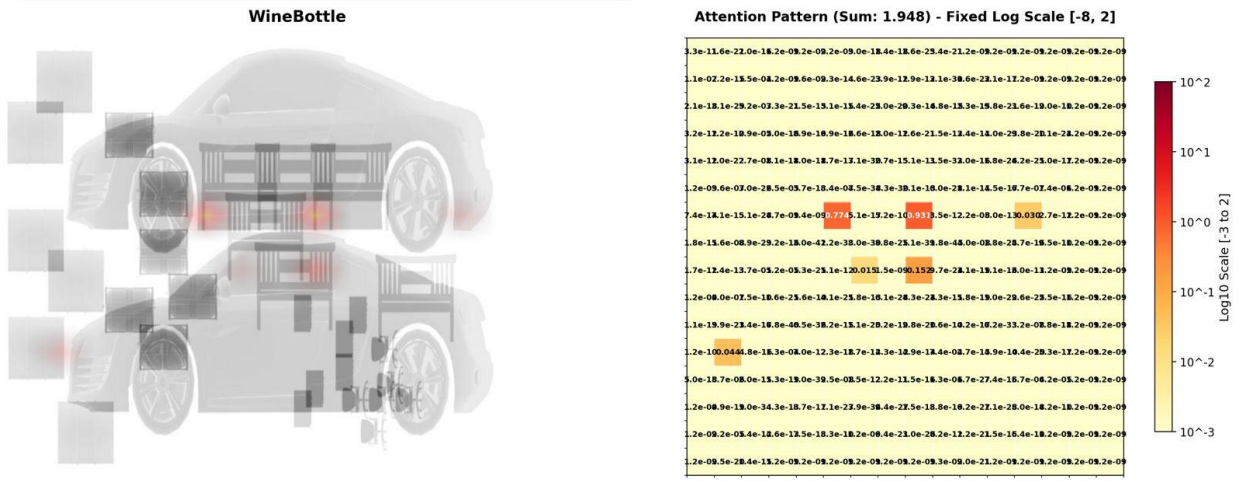


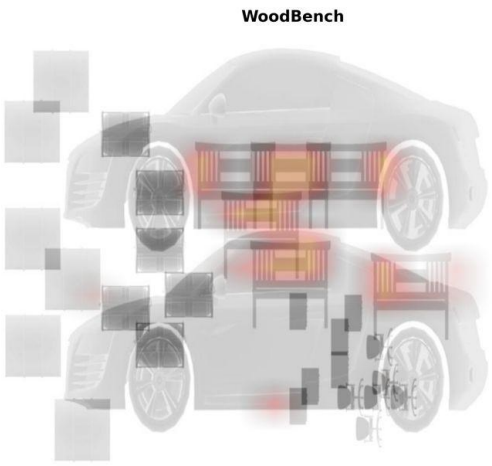
- Negotiated WSOL
- Happy path
  - Correct candidate is found
  - Wrong candidate isn’t

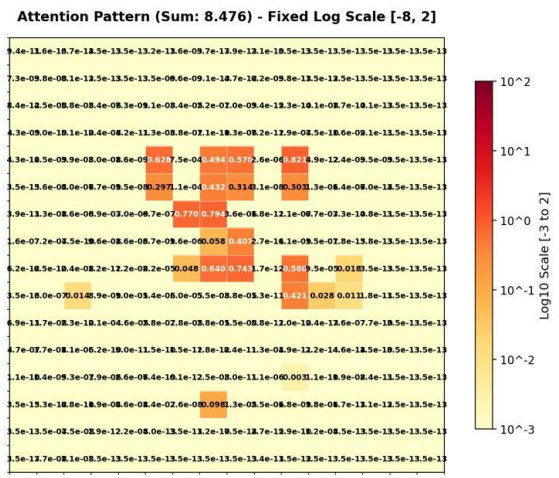
Negotiated path

- The correct candidate is found, but another candidate is mistakenly located elsewhere.
- However, the location of the wrong candidate is occupied by another, more clearly recognized item.
- The model seeks the best solution for all items simultaneously.



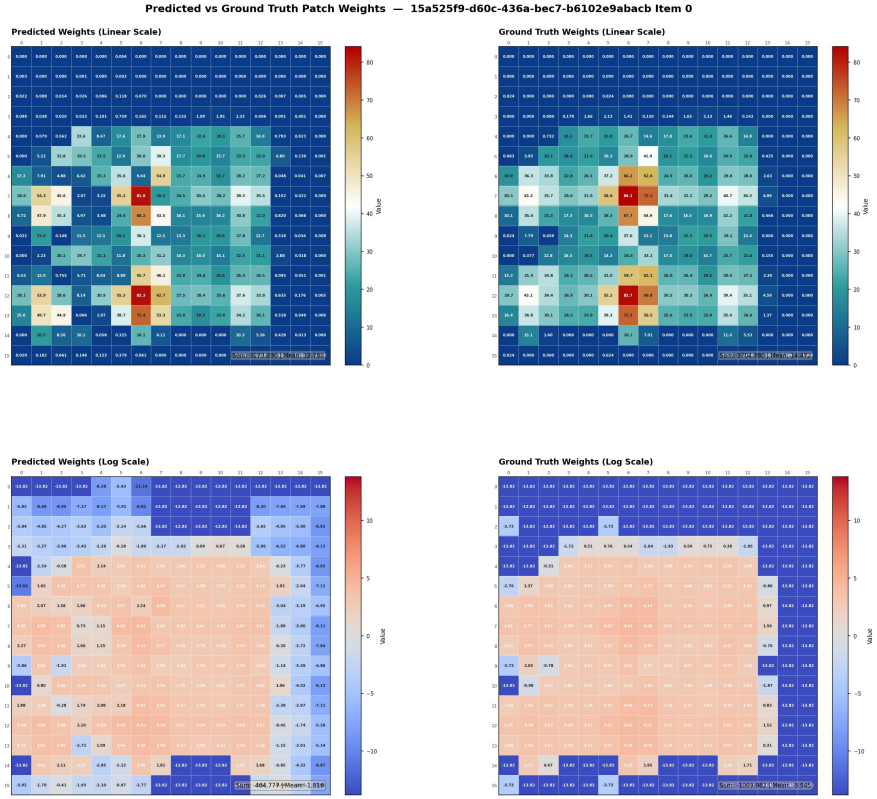


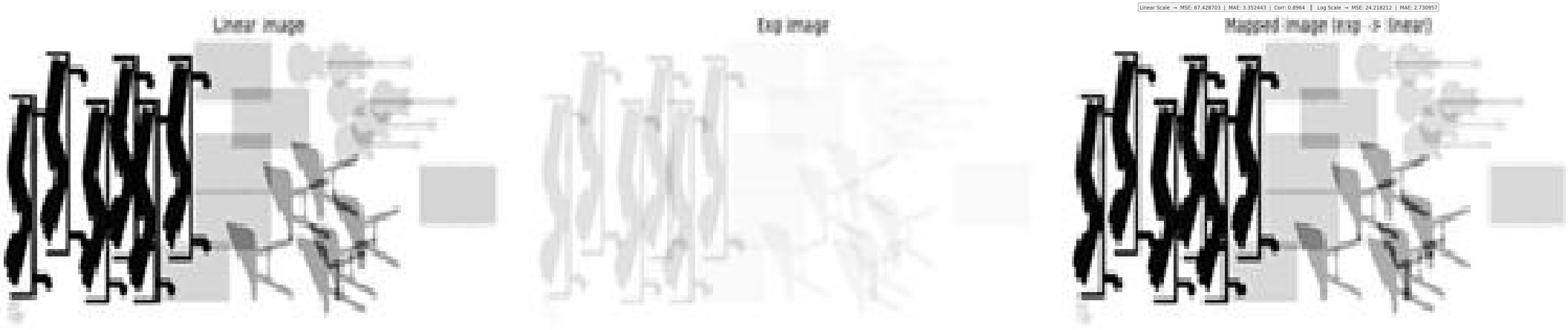




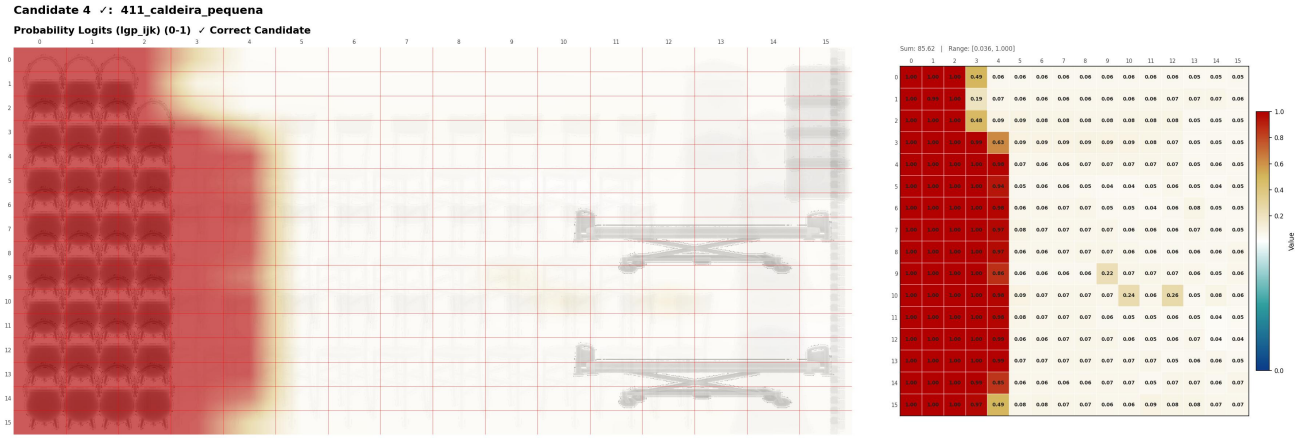
Auxiliary model for patch weights

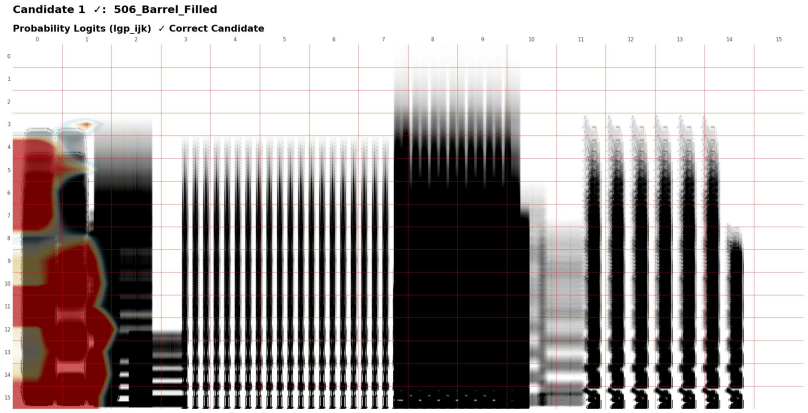
- Scanner calibration
- Second dimension for loss function od the main model

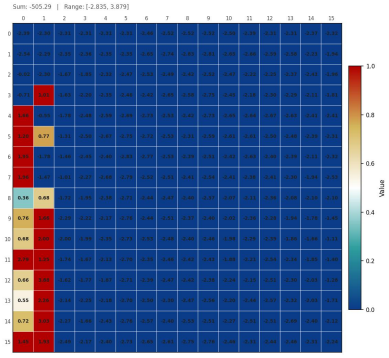


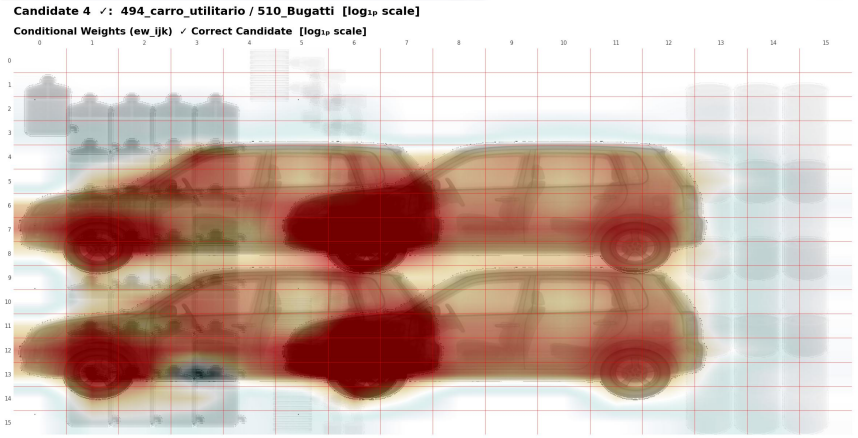


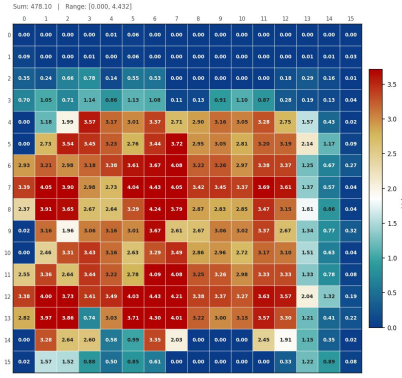
Some result maps

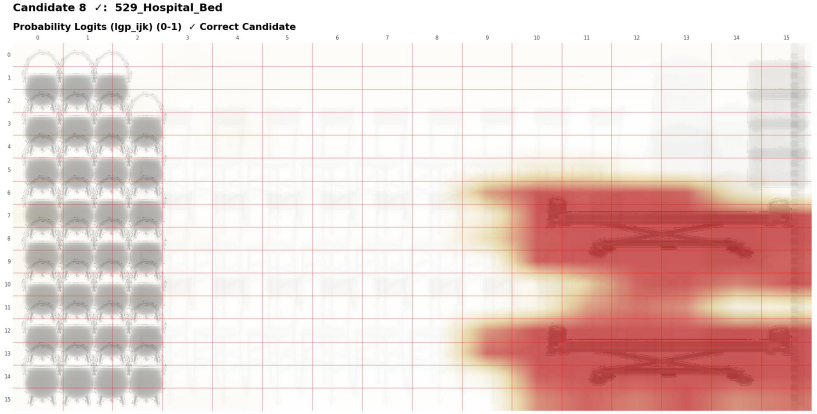


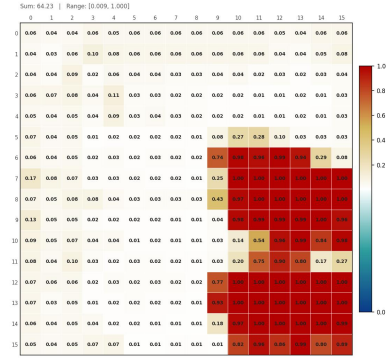










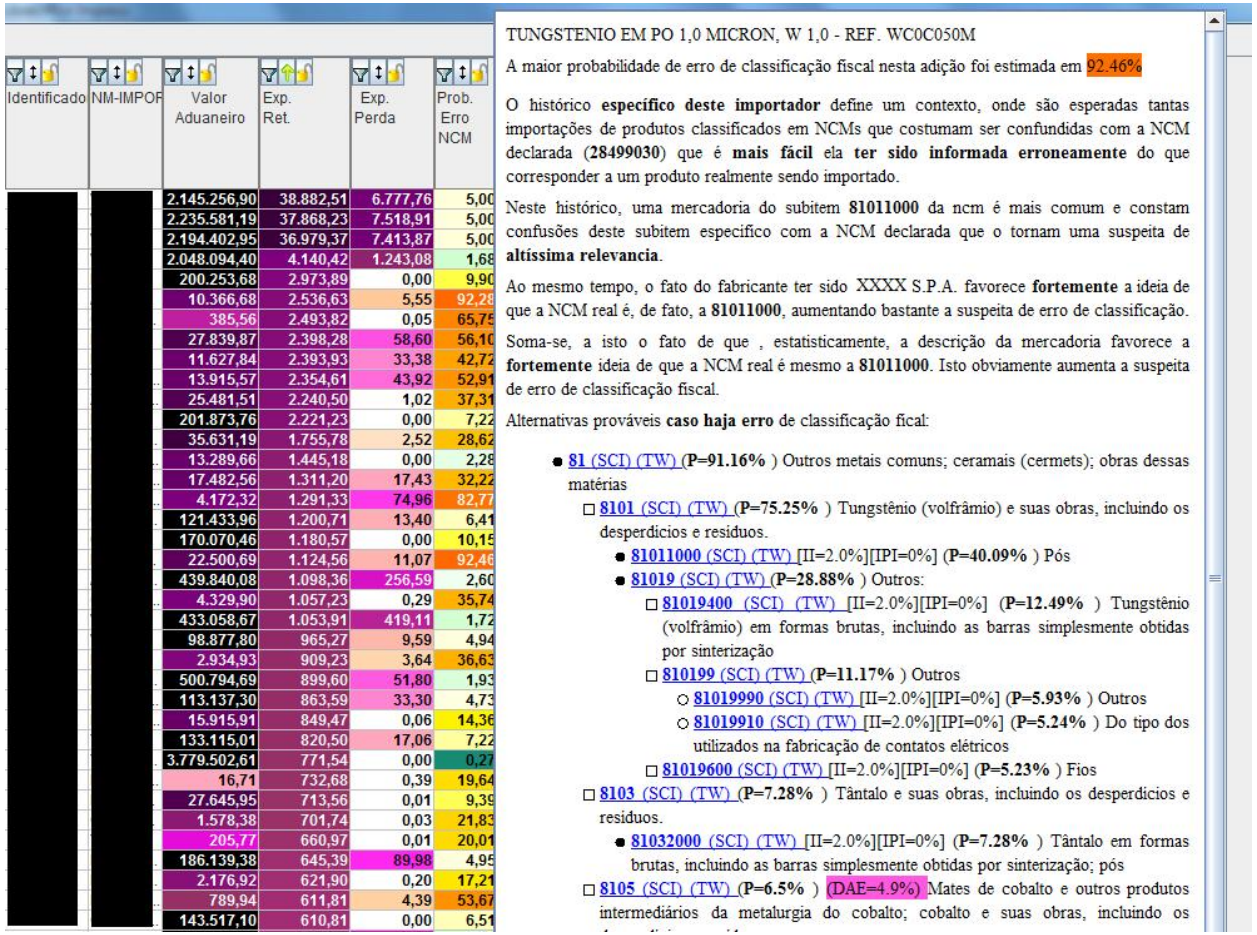


Handling items of the same tax code

- Sisam produces probability distributions for tax codes item by item
- X-Class produces probability distributions for the total weight goods of each tax code.
- We need the most reasonable predictions for each item considering both.
- They may rule out or reinforce different hypothesis about the items.
- We can find posterior probability distributions using MCMC simulation.

Expectations

- Not to present a new system to custom officers
- They should just notice na improvement in Sisam's results.
- Better probabilities
- Better justifications
  - Including images



Identificado	NM-IMPOR	Valor Aduaneiro	Exp. Ret.	Exp. Perda	Prob. Erro NCM
[redacted]	[redacted]	2.145.256,90	38.882,51	6.777,76	5,00
		2.235.581,19	37.868,23	7.518,91	5,00
		2.194.402,95	36.979,37	7.413,87	5,00
		2.048.094,40	4.140,42	1.243,08	1,68
		200.253,68	2.973,89	0,00	9,90
		10.366,68	2.536,63	5,55	92,28
		385,56	2.493,82	0,05	65,75
		27.839,87	2.398,28	58,60	56,10
		11.627,84	2.393,93	33,38	42,72
		13.915,57	2.354,61	43,92	52,91
		25.481,51	2.240,50	1,02	37,31
		201.873,76	2.221,23	0,00	7,22
		35.631,19	1.755,78	2,52	28,62
		13.289,66	1.445,18	0,00	2,28
		17.482,56	1.311,20	17,43	32,22
		4.172,32	1.291,33	74,96	82,77
		121.433,96	1.200,71	13,40	6,41
		170.070,46	1.180,57	0,00	10,15
		22.500,69	1.124,56	11,07	92,46
		439.840,08	1.098,36	256,59	2,60
		4.329,90	1.057,23	0,29	35,74
		433.058,67	1.053,91	419,11	1,72
		98.877,80	965,27	9,59	4,94
		2.934,93	909,23	3,64	36,63
		500.794,69	899,60	51,80	1,93
		113.137,30	863,59	33,30	4,73
		15.915,91	849,47	0,06	14,36
		133.115,01	820,50	17,06	7,22
		3.779.502,61	771,54	0,00	0,27
		16,71	732,68	0,39	19,64
		27.645,95	713,56	0,01	9,39
		1.578,38	701,74	0,03	21,83
		205,77	660,97	0,01	20,01
		186.139,38	645,39	89,98	4,95
		2.176,92	621,90	0,20	17,21
		789,94	611,81	4,39	53,67
		143.517,10	610,81	0,00	6,51

TUNGSTENIO EM PO 1,0 MICRON, W 1,0 - REF. WC0C050M

A maior probabilidade de erro de classificação fiscal nesta adição foi estimada em 92.46%

O histórico **específico deste importador** define um contexto, onde são esperadas tantas importações de produtos classificados em NCMs que costumam ser confundidas com a NCM declarada (**28499030**) que é **mais fácil** ela **ter sido informada erroneamente** do que corresponder a um produto realmente sendo importado.

Neste histórico, uma mercadoria do subitem **81011000** da ncm é mais comum e constam confusões deste subitem especifico com a NCM declarada que o tornam uma suspeita de **altíssima relevancia**.

Ao mesmo tempo, o fato do fabricante ter sido XXXX S.P.A. favorece **fortemente** a ideia de que a NCM real é, de fato, a **81011000**, aumentando bastante a suspeita de erro de classificação.

Soma-se, a isto o fato de que , estatisticamente, a descrição da mercadoria favorece a **fortemente** ideia de que a NCM real é mesmo a **81011000**. Isto obviamente aumenta a suspeita de erro de classificação fiscal.

Alternativas prováveis **caso haja erro** de classificação fical:

- 81 (SCI) (TW) (**P=91.16%**) Outros metais comuns; ceramais (cermets); obras dessas matérias
  - 8101 (SCI) (TW) (**P=75.25%**) Tungstênio (volfrâmio) e suas obras, incluindo os desperdicios e residuos.
    - 81011000 (SCI) (TW) [II=2.0%][IPI=0%] (**P=40.09%**) Pós
    - 81019 (SCI) (TW) (**P=28.88%**) Outros:
      - 81019400 (SCI) (TW) [II=2.0%][IPI=0%] (**P=12.49%**) Tungstênio (volfrâmio) em formas brutas, incluindo as barras simplesmente obtidas por sinterização
      - 81019910 (SCI) (TW) [II=2.0%][IPI=0%] (**P=5.24%**) Do tipo dos utilizados na fabricação de contatos elétricos
      - 81019990 (SCI) (TW) [II=2.0%][IPI=0%] (**P=5.93%**) Outros
    - 81019600 (SCI) (TW) [II=2.0%][IPI=0%] (**P=5.23%**) Fios
  - 8103 (SCI) (TW) (**P=7.28%**) Tântalo e suas obras, incluindo os desperdicios e residuos.
    - 81032000 (SCI) (TW) [II=2.0%][IPI=0%] (**P=7.28%**) Tântalo em formas brutas, incluindo as barras simplesmente obtidas por sinterização; pós
  - 8105 (SCI) (TW) (**P=6.5%**) (DAE=4.9%) Mates de cobalto e outros produtos intermediários da metalurgia do cobalto; cobalto e suas obras, incluindo os