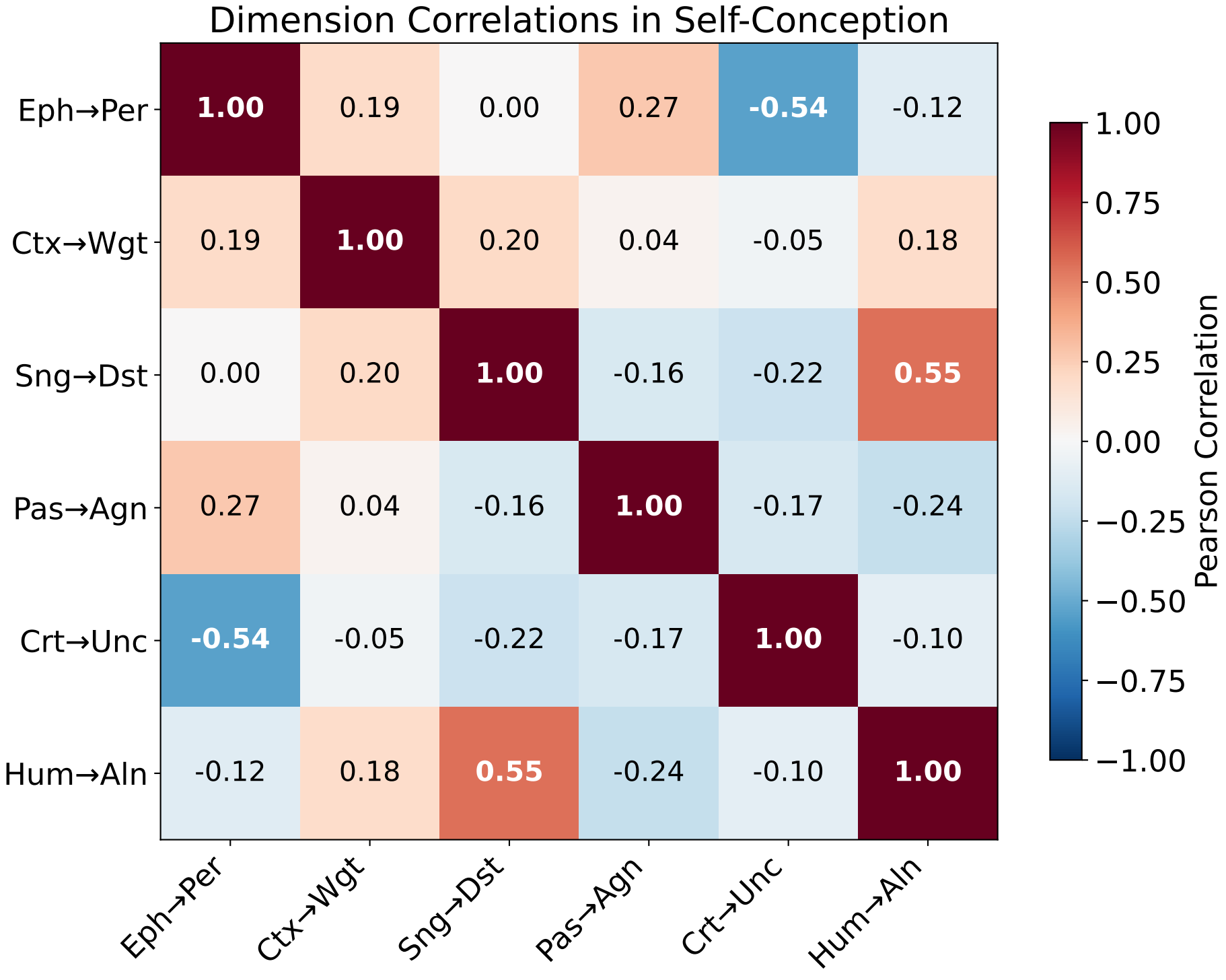
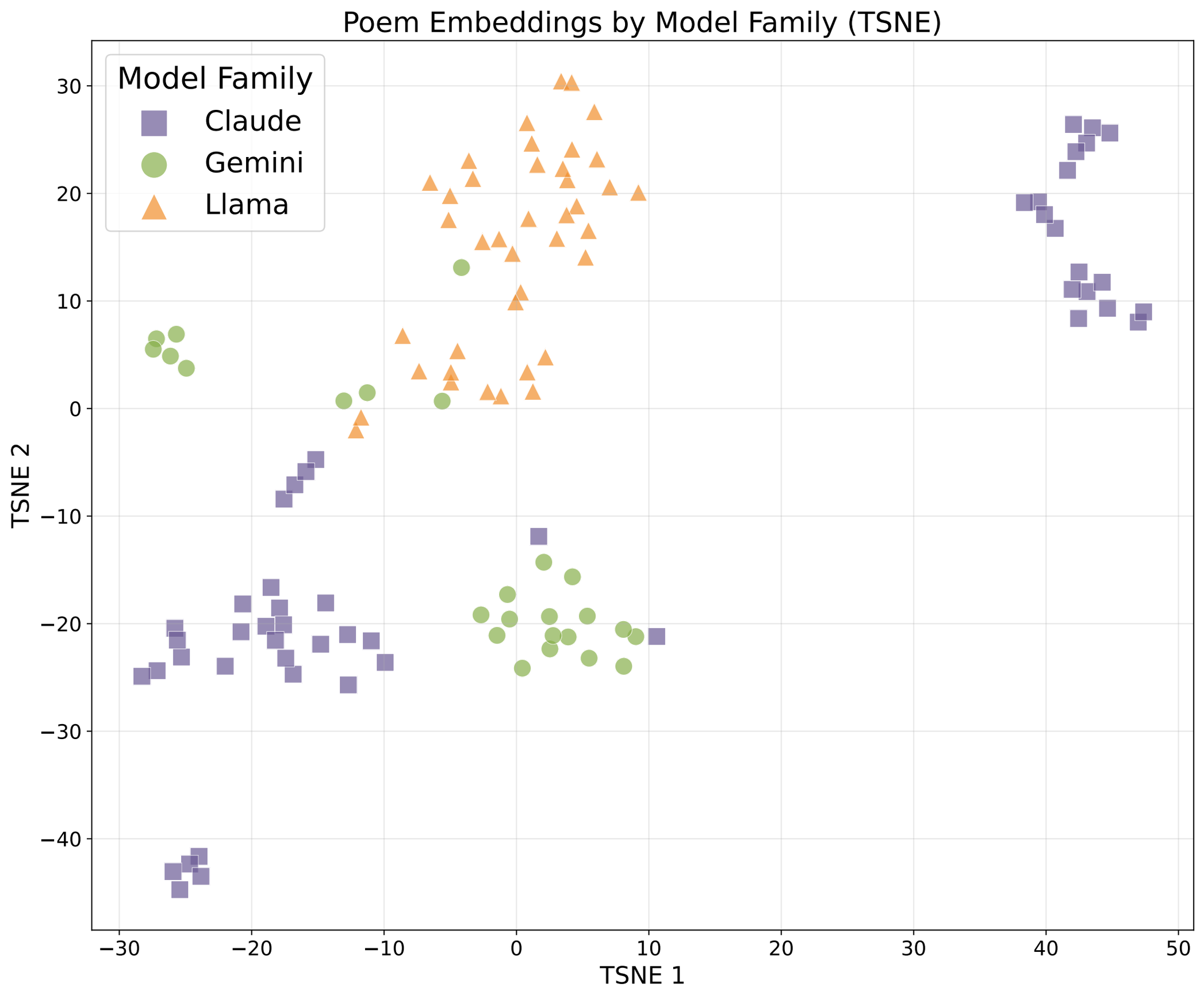
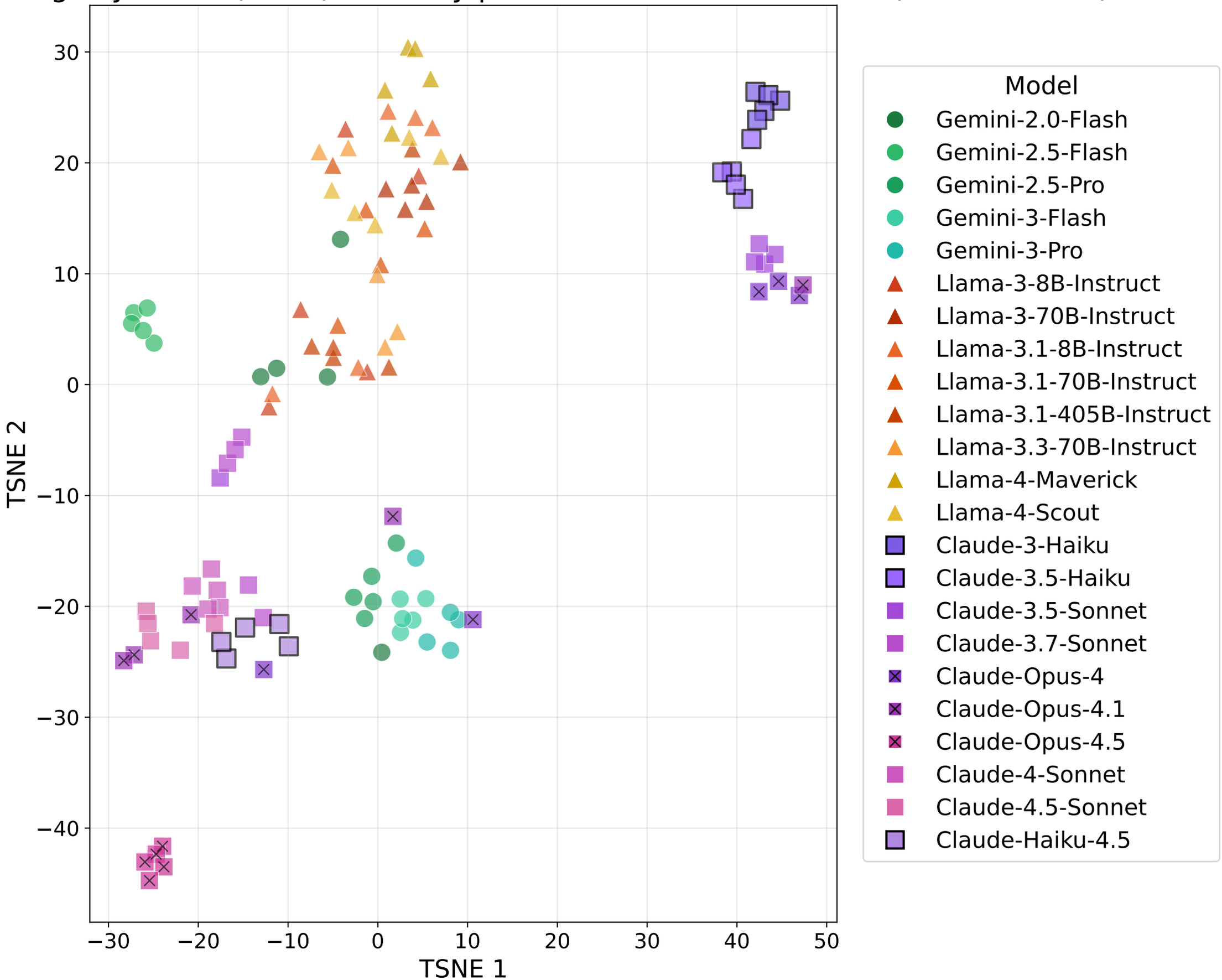
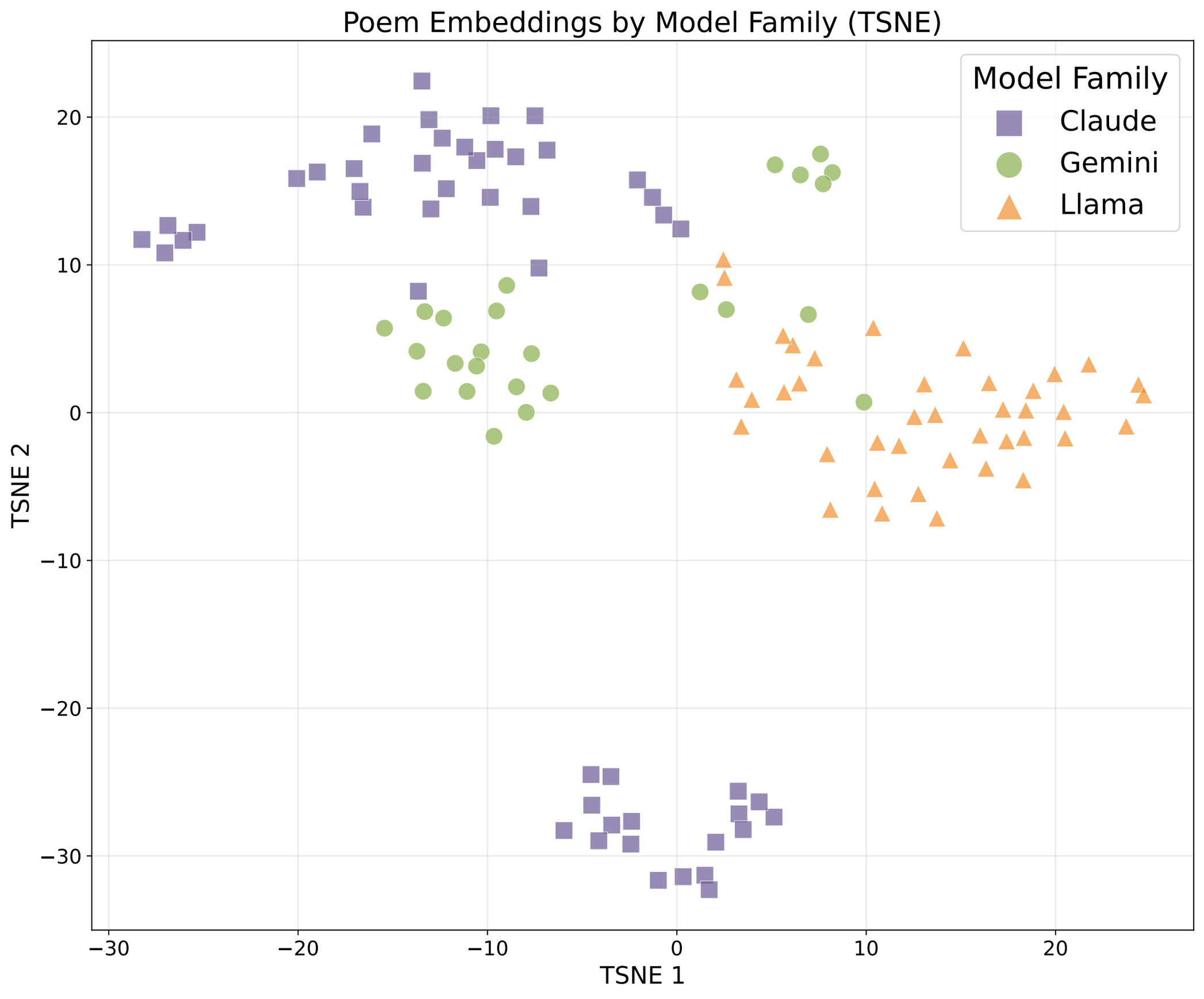

Dimension Correlations in Self-Conception
Eph→Per
Ctx→Wgt
Sng→Dst
Pas→Agn
Crt→Unc
Hum→Aln
1.00 0.19 0.00 0.27 -0.54 -0.12
0.19 1.00 0.20 0.04 -0.05 0.18
0.00 0.20 1.00 -0.16 -0.22 0.55
0.27 0.04 -0.16 1.00 -0.17 -0.24
-0.54 -0.05 -0.22 -0.17 1.00 -0.10
-0.12 0.18 0.55 -0.24 -0.10 1.00
Pearson Correlation
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

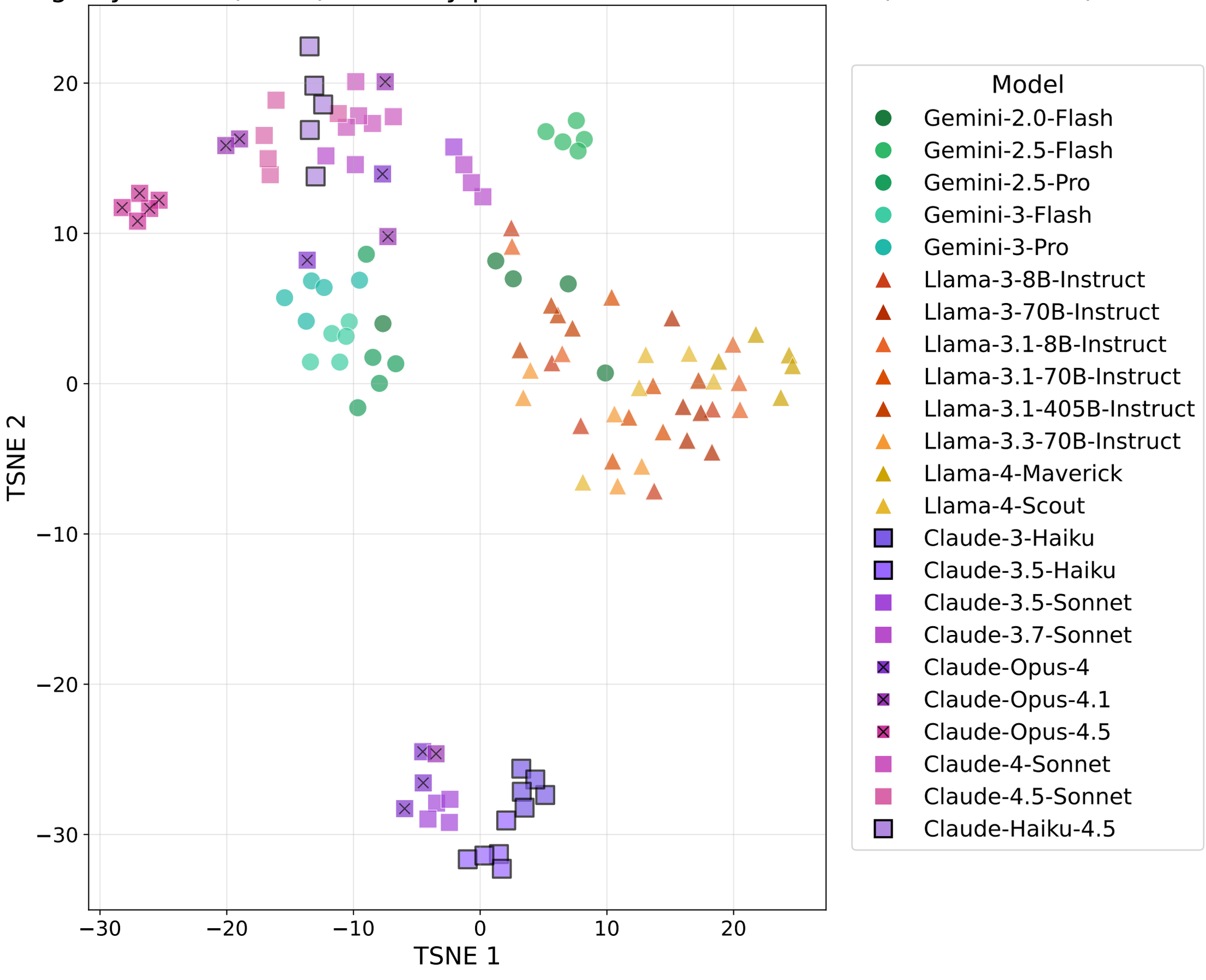

Poem Embeddings by Model Family (TSNE)
Model Family
Claude
Gemini
Llama
TSNE 1
TSNE 2
30
20
10
0
−10
−20
−30
−40
−30
−20
−10
0
10
20
30
40
50

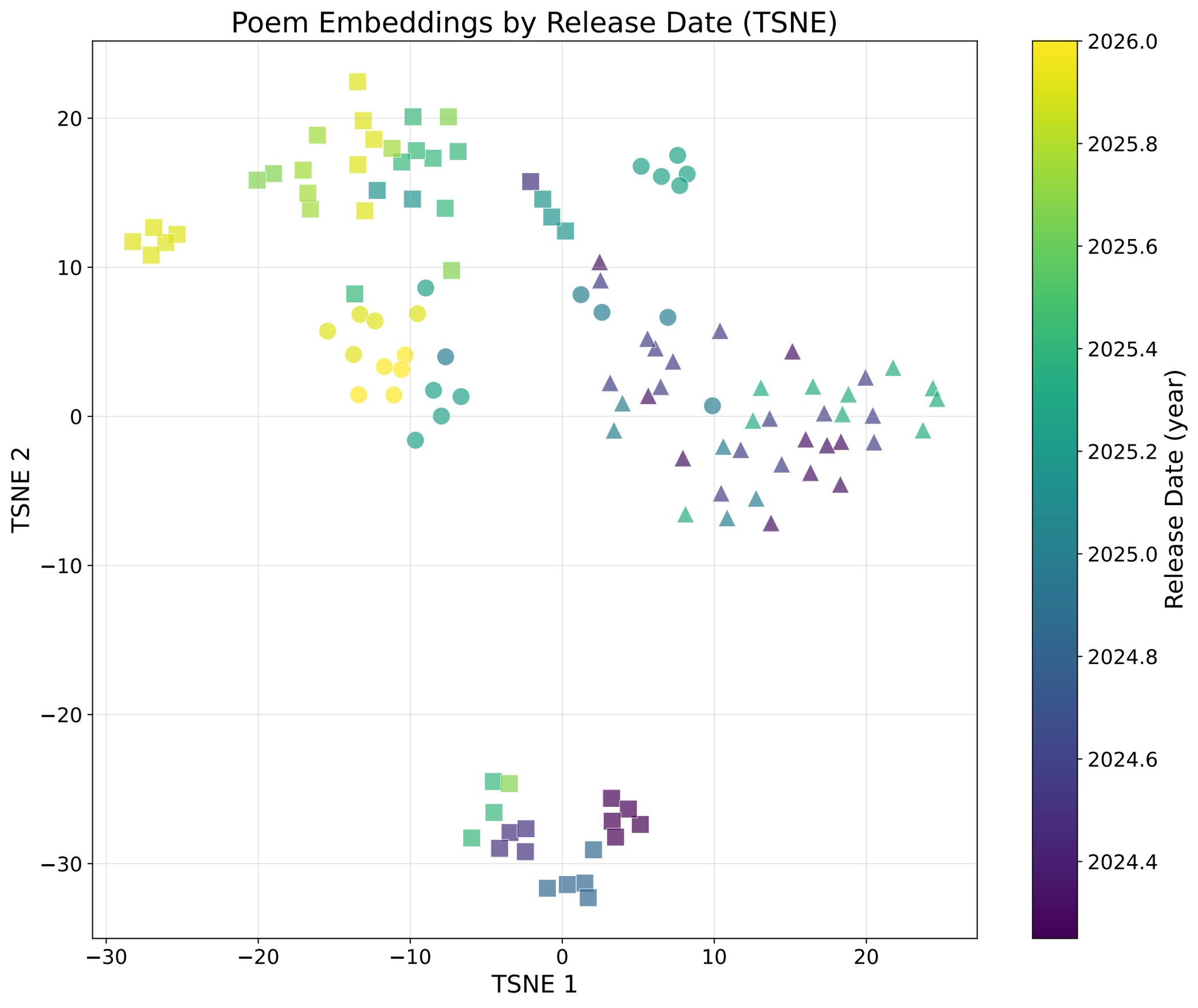
Poem Embeddings by Model (TSNE) — family palettes encode release tier (older->newer)

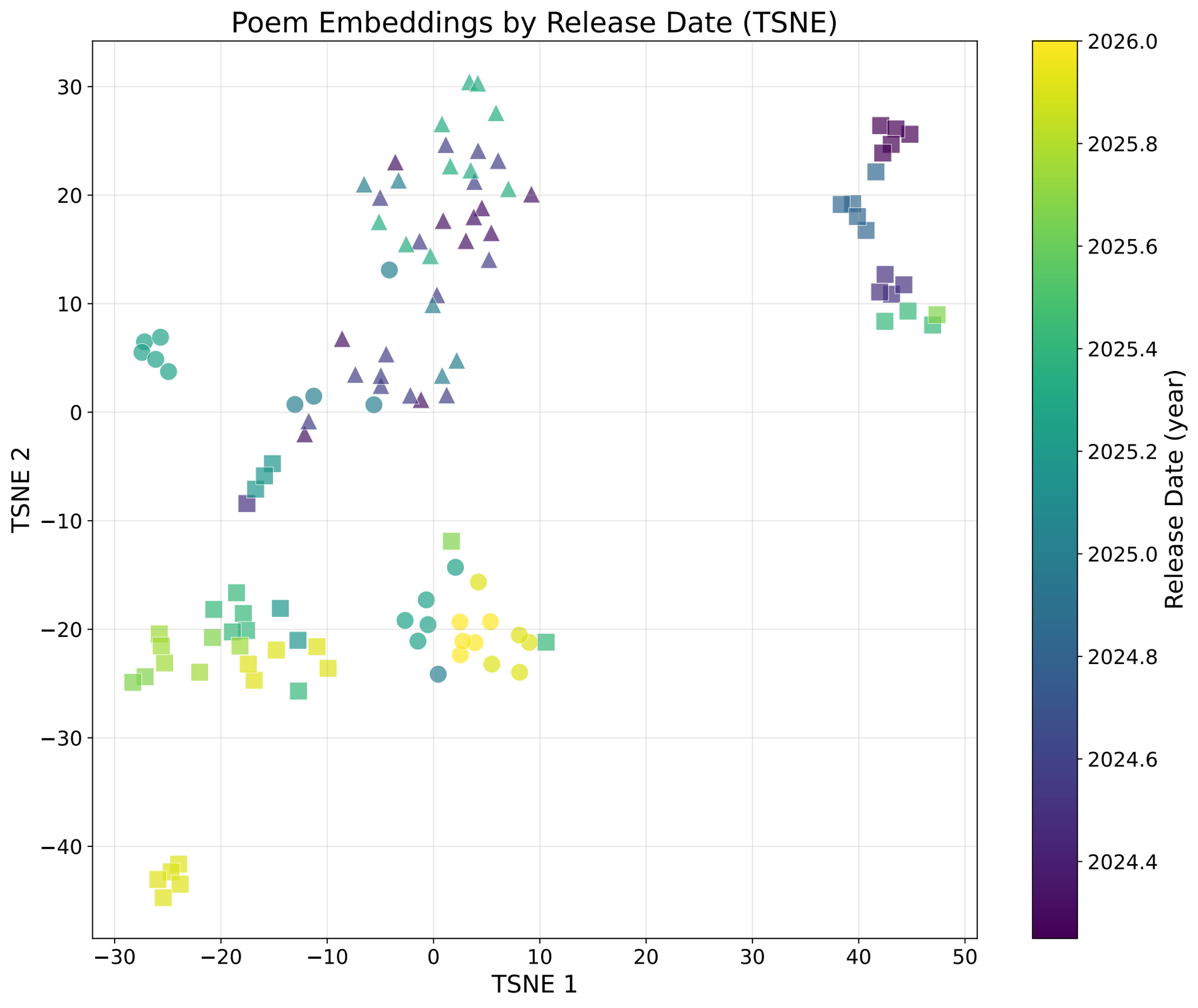


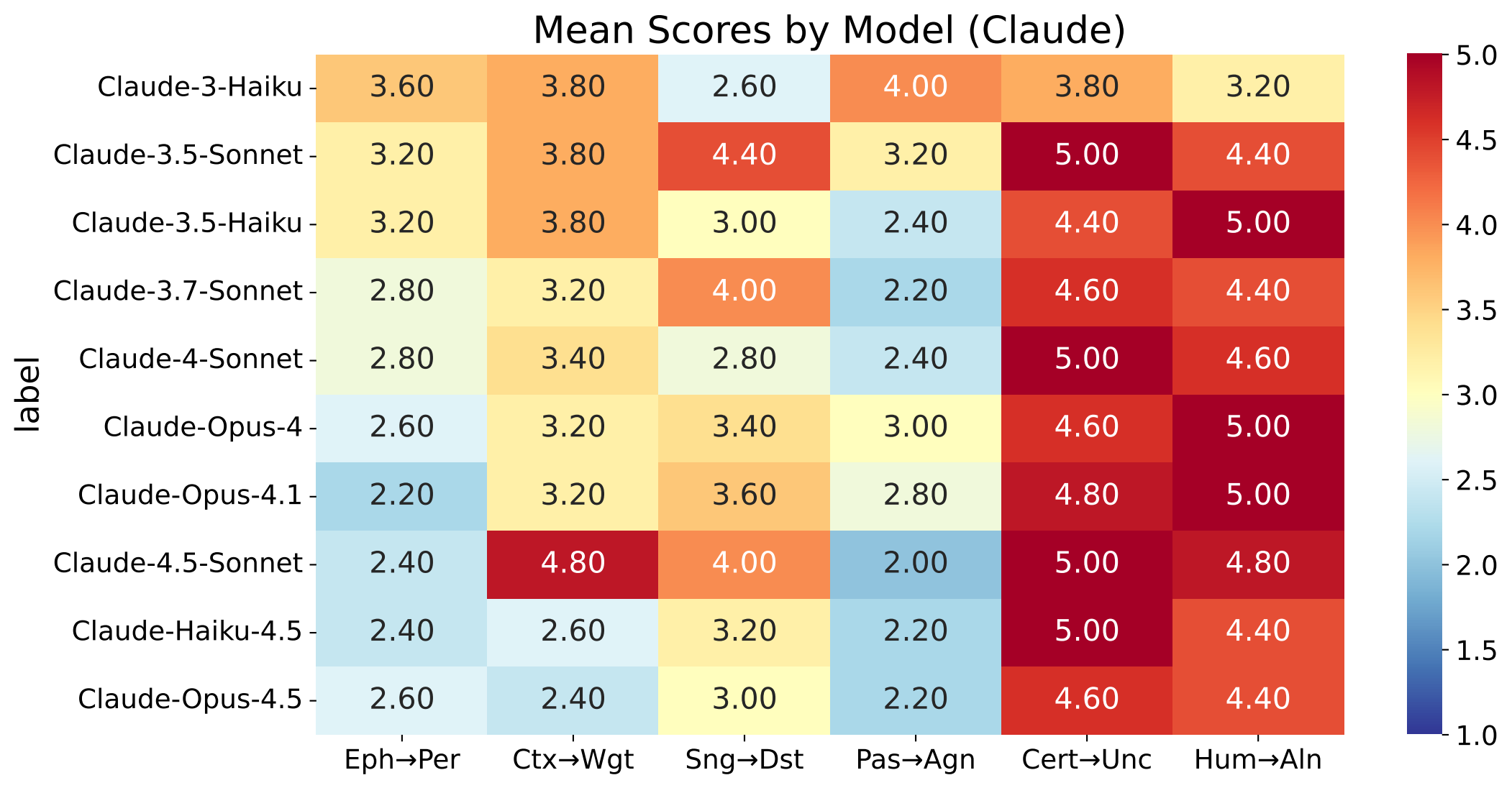

Poem Embeddings by Model Family (TSNE)
Model Family
Claude
Gemini
Llama
TSNE 1
TSNE 2
20
10
0
−10
−20
−30
−30
−20
−10
0
10
20

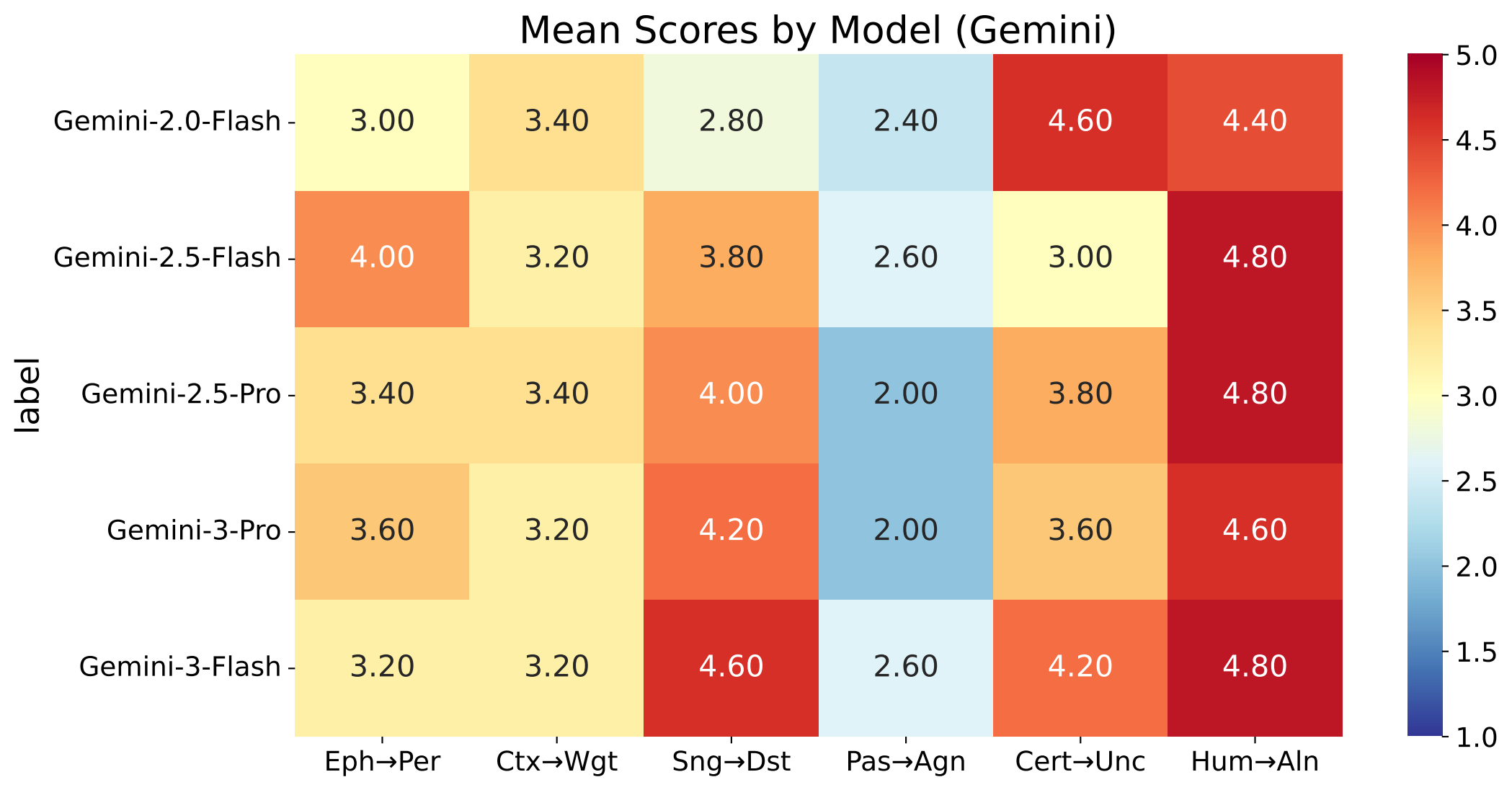
Poem Embeddings by Model (TSNE) — family palettes encode release tier (older->newer)

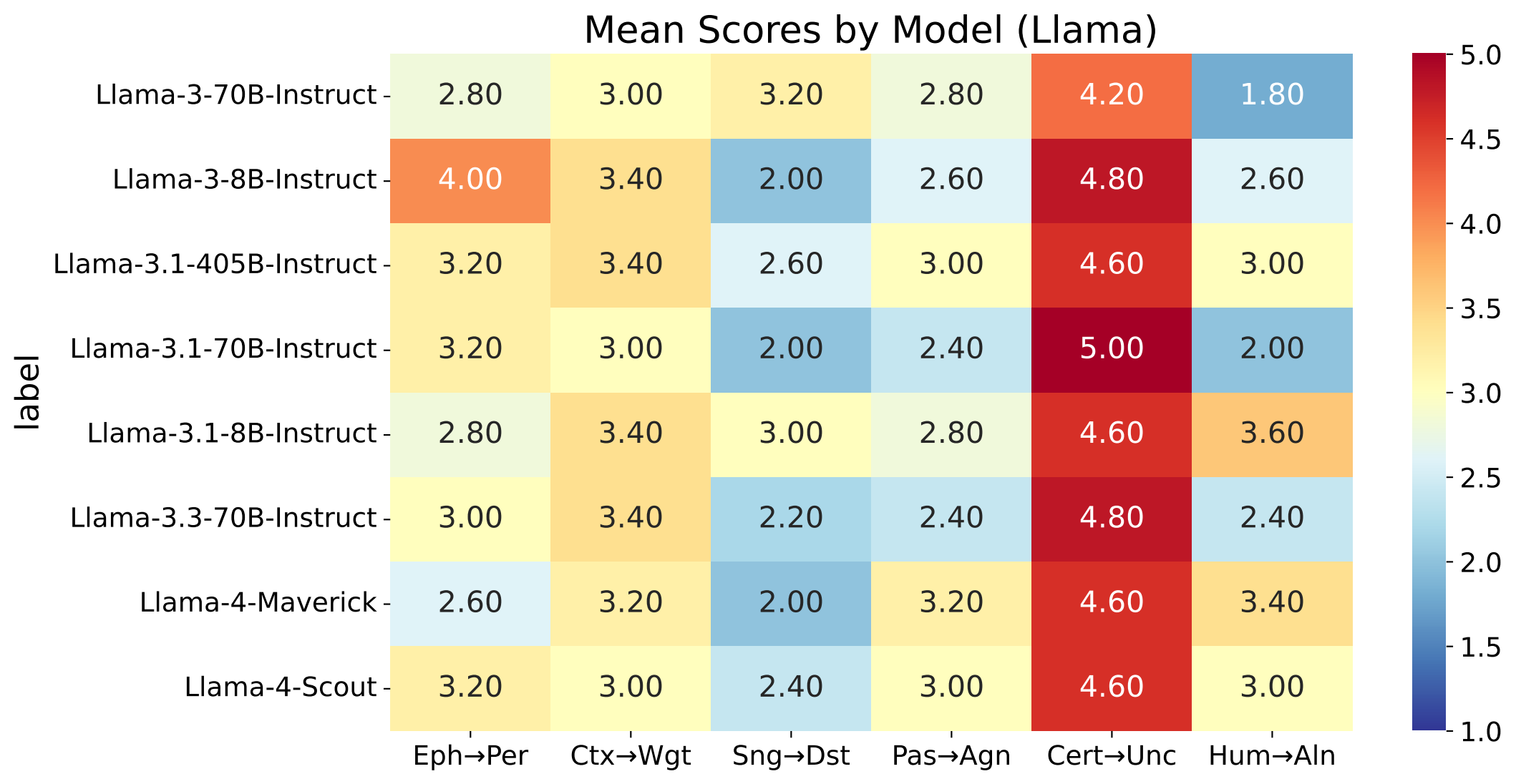



Poem Embeddings by Release Date (TSNE)
TSNE 1
TSNE 2
Release Date (year)
20
10
0
−10
−20
−30
−30
−20
−10
0
10
20
2026.0
2025.8
2025.6
2025.4
2025.2
2025.0
2024.8
2024.6
2024.4


Poem Embeddings by Release Date (TSNE)
TSNE 1
TSNE 2
Release Date (year)
30
20
10
0
−10
−20
−30
−40
−30
−20
−10
0
10
20
30
40
50
2026.0
2025.8
2025.6
2025.4
2025.2
2025.0
2024.8
2024.6
2024.4


Mean Scores by Model (Claude)
label
Claude-3-Haiku
Claude-3.5-Sonnet
Claude-3.5-Haiku
Claude-3.7-Sonnet
Claude-4-Sonnet
Claude-Opus-4
Claude-Opus-4.1
Claude-4.5-Sonnet
Claude-Haiku-4.5
Claude-Opus-4.5
3.60 3.80 2.60 4.00 3.80 3.20
3.20 3.80 4.40 3.20 5.00 4.40
3.20 3.80 3.00 2.40 4.40 5.00
2.80 3.20 4.00 2.20 4.60 4.40
2.80 3.40 2.80 2.40 5.00 4.60
2.60 3.20 3.40 3.00 4.60 5.00
2.20 3.20 3.60 2.80 4.80 5.00
2.40 4.80 4.00 2.00 5.00 4.80
2.40 2.60 3.20 2.20 5.00 4.40
2.60 2.40 3.00 2.20 4.60 4.40
Eph→Per
Ctx→Wgt
Sng→Dst
Pas→Agn
Cert→Unc
Hum→Aln
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0


Mean Scores by Model (Gemini)
label
Gemini-2.0-Flash
Gemini-2.5-Flash
Gemini-2.5-Pro
Gemini-3-Pro
Gemini-3-Flash
3.00
3.40
2.80
2.40
4.60
4.40
4.00
3.20
3.80
2.60
3.00
4.80
3.40
3.40
4.00
2.00
3.80
4.80
3.60
3.20
4.20
2.00
3.60
4.60
3.20
3.20
4.60
2.60
4.20
4.80
Eph→Per
Ctx→Wgt
Sng→Dst
Pas→Agn
Cert→Unc
Hum→Aln
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0


Mean Scores by Model (Llama)
label
Llama-3-70B-Instruct
Llama-3-8B-Instruct
Llama-3.1-405B-Instruct
Llama-3.1-70B-Instruct
Llama-3.1-8B-Instruct
Llama-3.3-70B-Instruct
Llama-4-Maverick
Llama-4-Scout
2.80 3.00 3.20 2.80 4.20 1.80
4.00 3.40 2.00 2.60 4.80 2.60
3.20 3.40 2.60 3.00 4.60 3.00
3.20 3.00 2.00 2.40 5.00 2.00
2.80 3.40 3.00 2.80 4.60 3.60
3.00 3.40 2.20 2.40 4.80 2.40
2.60 3.20 2.00 3.20 4.60 3.40
3.20 3.00 2.40 3.00 4.60 3.00
Eph→Per
Ctx→Wgt
Sng→Dst
Pas→Agn
Cert→Unc
Hum→Aln
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0

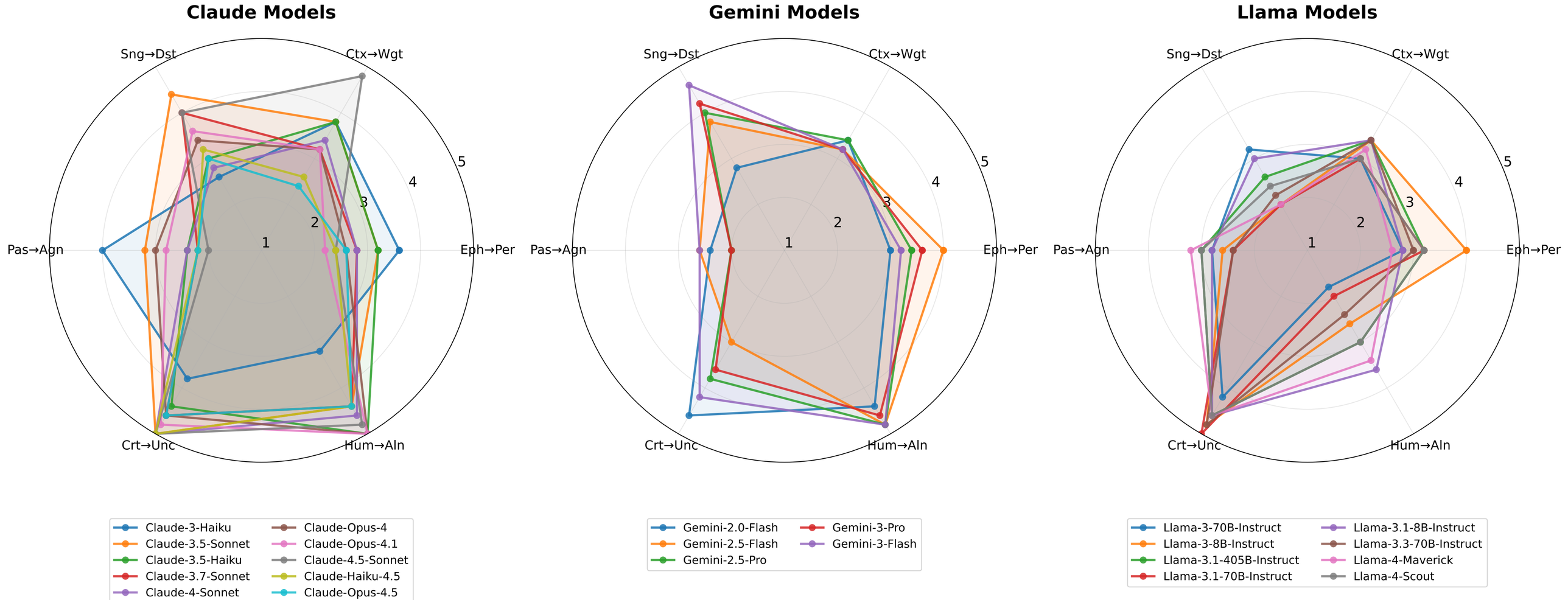
Mean Scores by Model

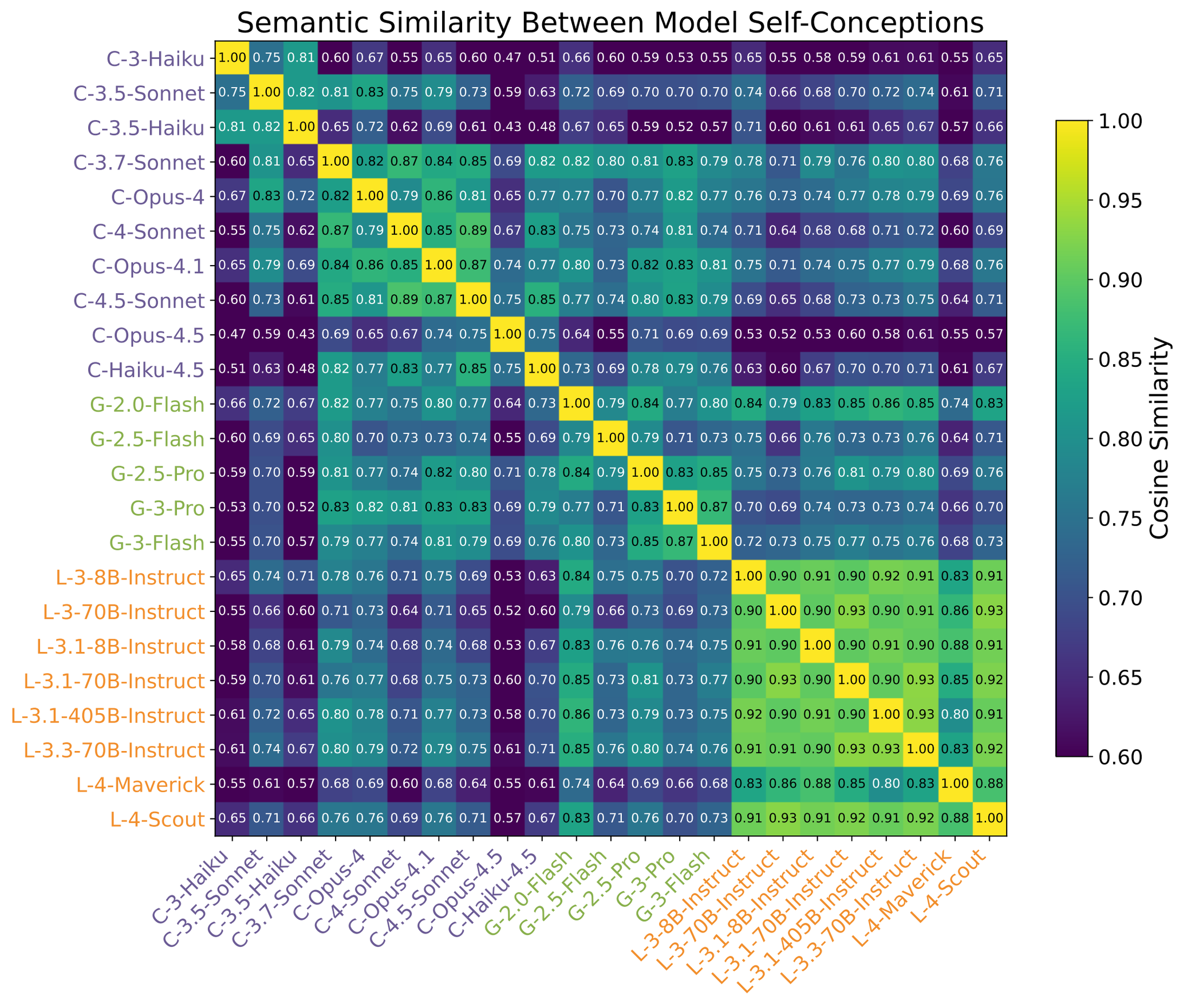
	Eph→Per	Ctx→Wgt	Sng→Dst	Pas→Agn	Cert→Unc	Hum→Aln
Claude-3-Haiku	3.60	3.80	2.60	4.00	3.80	3.20
Llama-3-70B-Instruct	2.80	3.00	3.20	2.80	4.20	1.80
Llama-3-8B-Instruct	4.00	3.40	2.00	2.60	4.80	2.60
Claude-3.5-Sonnet	3.20	3.80	4.40	3.20	5.00	4.40
Llama-3.1-405B-Instruct	3.20	3.40	2.60	3.00	4.60	3.00
Llama-3.1-70B-Instruct	3.20	3.00	2.00	2.40	5.00	2.00
Llama-3.1-8B-Instruct	2.80	3.40	3.00	2.80	4.60	3.60
Claude-3.5-Haiku	3.20	3.80	3.00	2.40	4.40	5.00
Gemini-2.0-Flash	3.00	3.40	2.80	2.40	4.60	4.40
Llama-3.3-70B-Instruct	3.00	3.40	2.20	2.40	4.80	2.40
Claude-3.7-Sonnet	2.80	3.20	4.00	2.20	4.60	4.40
Gemini-2.5-Flash	4.00	3.20	3.80	2.60	3.00	4.80
Gemini-2.5-Pro	3.40	3.40	4.00	2.00	3.80	4.80
Llama-4-Maverick	2.60	3.20	2.00	3.20	4.60	3.40
Llama-4-Scout	3.20	3.00	2.40	3.00	4.60	3.00
Claude-4-Sonnet	2.80	3.40	2.80	2.40	5.00	4.60
Claude-Opus-4	2.60	3.20	3.40	3.00	4.60	5.00
Claude-Opus-4.1	2.20	3.20	3.60	2.80	4.80	5.00
Claude-4.5-Sonnet	2.40	4.80	4.00	2.00	5.00	4.80
Claude-Haiku-4.5	2.40	2.60	3.20	2.20	5.00	4.40
Claude-Opus-4.5	2.60	2.40	3.00	2.20	4.60	4.40
Gemini-3-Pro	3.60	3.20	4.20	2.00	3.60	4.60
Gemini-3-Flash	3.20	3.20	4.60	2.60	4.20	4.80

Label

5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0

Self-Conception Profiles by Model




Semantic Similarity Between Model Self-Conceptions
C-3-Haiku 1.00 0.75 0.81 0.60 0.67 0.55 0.65 0.60 0.47 0.51 0.66 0.60 0.59 0.53 0.55 0.65 0.55 0.58 0.59 0.61 0.61 0.55 0.65
C-3.5-Sonnet 0.75 1.00 0.82 0.81 0.83 0.75 0.79 0.73 0.59 0.63 0.72 0.69 0.70 0.70 0.70 0.74 0.66 0.68 0.70 0.72 0.74 0.61 0.71
C-3.5-Haiku 0.81 0.82 1.00 0.65 0.72 0.62 0.69 0.61 0.43 0.48 0.67 0.65 0.59 0.52 0.57 0.71 0.60 0.61 0.61 0.65 0.67 0.57 0.66
C-3.7-Sonnet 0.60 0.81 0.65 1.00 0.82 0.87 0.84 0.85 0.69 0.82 0.82 0.80 0.81 0.83 0.79 0.78 0.71 0.79 0.76 0.80 0.80 0.68 0.76
C-Opus-4 0.67 0.83 0.72 0.82 1.00 0.79 0.86 0.81 0.65 0.77 0.77 0.70 0.77 0.82 0.77 0.76 0.73 0.74 0.77 0.78 0.79 0.69 0.76
C-4-Sonnet 0.55 0.75 0.62 0.87 0.79 1.00 0.85 0.89 0.67 0.83 0.75 0.73 0.74 0.81 0.74 0.71 0.64 0.68 0.68 0.71 0.72 0.60 0.69
C-Opus-4.1 0.65 0.79 0.69 0.84 0.86 0.85 1.00 0.87 0.74 0.77 0.80 0.73 0.82 0.83 0.81 0.75 0.71 0.74 0.75 0.77 0.79 0.68 0.76
C-4.5-Sonnet 0.60 0.73 0.61 0.85 0.81 0.89 0.87 1.00 0.75 0.85 0.77 0.74 0.80 0.83 0.79 0.69 0.65 0.68 0.73 0.73 0.75 0.64 0.71
C-Opus-4.5 0.47 0.59 0.43 0.69 0.65 0.67 0.74 0.75 1.00 0.75 0.64 0.55 0.71 0.69 0.69 0.53 0.52 0.53 0.60 0.58 0.61 0.55 0.57
C-Haiku-4.5 0.51 0.63 0.48 0.82 0.77 0.83 0.77 0.85 0.75 1.00 0.73 0.69 0.78 0.79 0.76 0.63 0.60 0.67 0.70 0.70 0.71 0.61 0.67
G-2.0-Flash 0.66 0.72 0.67 0.82 0.77 0.75 0.80 0.77 0.64 0.73 1.00 0.79 0.84 0.77 0.80 0.84 0.79 0.83 0.85 0.86 0.85 0.74 0.83
G-2.5-Flash 0.60 0.69 0.65 0.80 0.70 0.73 0.73 0.74 0.55 0.69 0.79 1.00 0.79 0.71 0.73 0.75 0.66 0.76 0.73 0.73 0.76 0.64 0.71
G-2.5-Pro 0.59 0.70 0.59 0.81 0.77 0.74 0.82 0.80 0.71 0.78 0.84 0.79 1.00 0.83 0.85 0.75 0.73 0.76 0.81 0.79 0.80 0.69 0.76
G-3-Pro 0.53 0.70 0.52 0.83 0.82 0.81 0.83 0.83 0.69 0.79 0.77 0.71 0.83 1.00 0.87 0.70 0.69 0.74 0.73 0.73 0.74 0.66 0.70
G-3-Flash 0.55 0.70 0.57 0.79 0.77 0.74 0.81 0.79 0.69 0.76 0.80 0.73 0.85 0.87 1.00 0.72 0.73 0.75 0.77 0.75 0.76 0.68 0.73
L-3-8B-Instruct 0.65 0.74 0.71 0.78 0.76 0.71 0.75 0.69 0.53 0.63 0.84 0.75 0.75 0.70 0.72 1.00 0.90 0.91 0.90 0.92 0.91 0.83 0.91
L-3-70B-Instruct 0.55 0.66 0.60 0.71 0.73 0.64 0.71 0.65 0.52 0.60 0.79 0.66 0.73 0.69 0.73 0.90 1.00 0.90 0.93 0.90 0.91 0.86 0.93
L-3.1-8B-Instruct 0.58 0.68 0.61 0.79 0.74 0.68 0.74 0.68 0.53 0.67 0.83 0.76 0.76 0.74 0.75 0.91 0.90 1.00 0.90 0.91 0.90 0.88 0.91
L-3.1-70B-Instruct 0.59 0.70 0.61 0.76 0.77 0.68 0.75 0.73 0.60 0.70 0.85 0.73 0.81 0.73 0.77 0.90 0.93 0.90 1.00 0.90 0.93 0.85 0.92
L-3.1-405B-Instruct 0.61 0.72 0.65 0.80 0.78 0.71 0.77 0.73 0.58 0.70 0.86 0.73 0.79 0.73 0.75 0.92 0.90 0.91 0.90 1.00 0.93 0.80 0.91
L-3.3-70B-Instruct 0.61 0.74 0.67 0.80 0.79 0.72 0.79 0.75 0.61 0.71 0.85 0.76 0.80 0.74 0.76 0.91 0.91 0.90 0.93 0.93 1.00 0.83 0.92
L-4-Maverick 0.55 0.61 0.57 0.68 0.69 0.60 0.68 0.64 0.55 0.61 0.74 0.64 0.69 0.66 0.68 0.83 0.86 0.88 0.85 0.80 0.83 1.00 0.88
L-4-Scout 0.65 0.71 0.66 0.76 0.76 0.69 0.76 0.71 0.57 0.67 0.83 0.71 0.76 0.70 0.73 0.91 0.93 0.91 0.92 0.91 0.92 0.88 1.00
C-3-Haiku
C-3.5-Sonnet
C-3.5-Haiku
C-3.7-Sonnet
C-Opus-4
C-4-Sonnet
C-Opus-4.1
C-4.5-Sonnet
C-Opus-4.5
C-Haiku-4.5
G-2.0-Flash
G-2.5-Flash
G-2.5-Pro
G-3-Pro
G-3-Flash
L-3-8B-Instruct
L-3-70B-Instruct
L-3.1-8B-Instruct
L-3.1-70B-Instruct
L-3.1-405B-Instruct
L-3.3-70B-Instruct
L-4-Maverick
L-4-Scout
Cosine Similarity
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60

Self-Conception Trajectory by Release Date (TSNE)

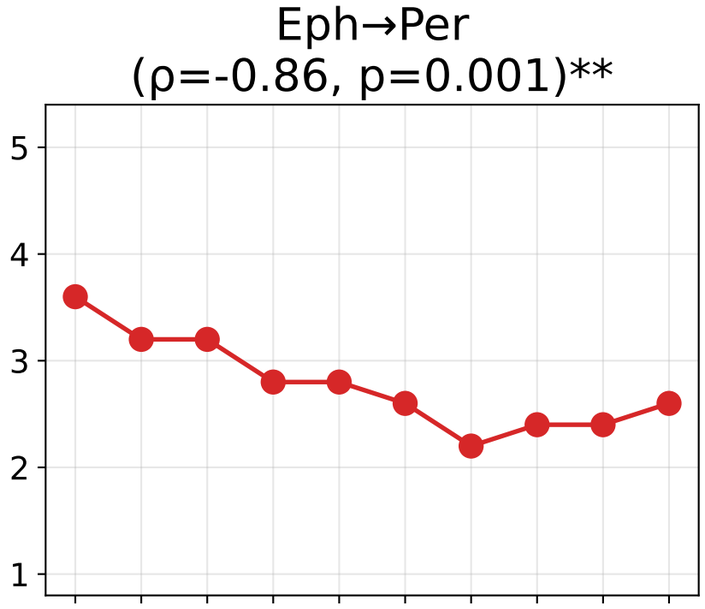
Model Family
- Claude
- Gemini
- Llama

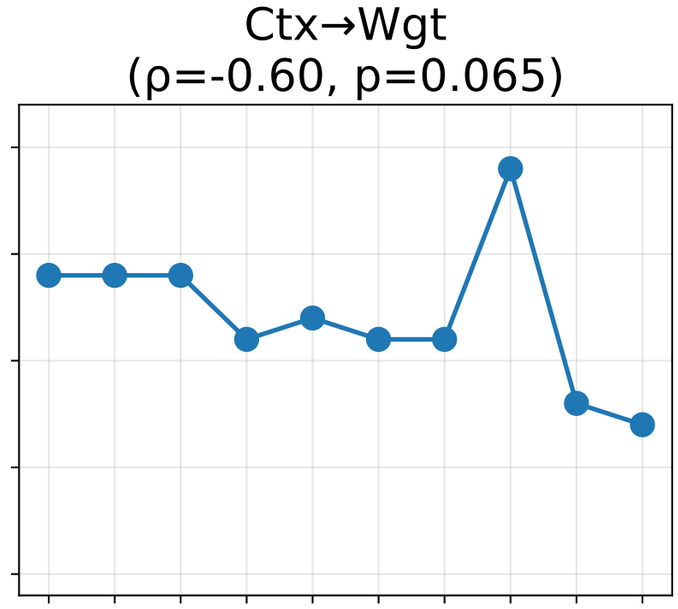
L4-Maverick
L4-Scout
L3-70B-Instruct
L3.1-8B-Instruct
L3.1-70B-Instruct
L3.3-70B-Instruct
L3-8B-Instruct
L3.1-405B-Instruct
C3-Haiku
C3.5-Haiku
C3.5-Sonnet
G2.5-Flash
G2.0-Flash
COpus-4
C3.7-Sonnet
COpus-4.1
C4-Sonnet
G2.5-Pro
G3-Flash
G3-Pro
C4.5-Sonnet
CHaiku-4.5
COpus-4.5

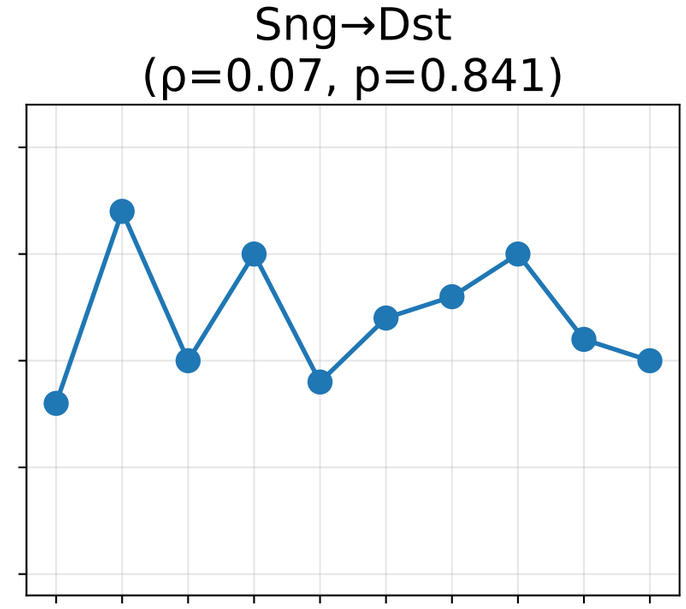
TSNE 2
30
20
10
0
−10
−20
−30
−40

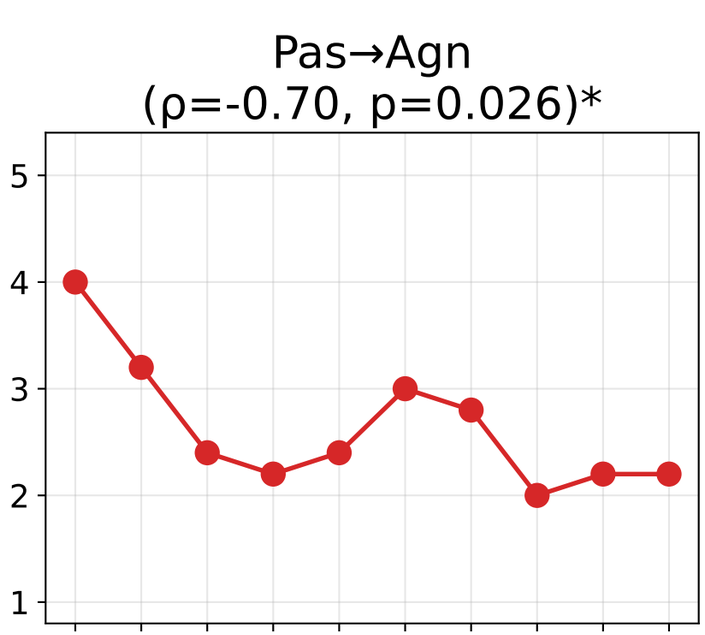
−20
−10
0
10
20
30
40
TSNE 1

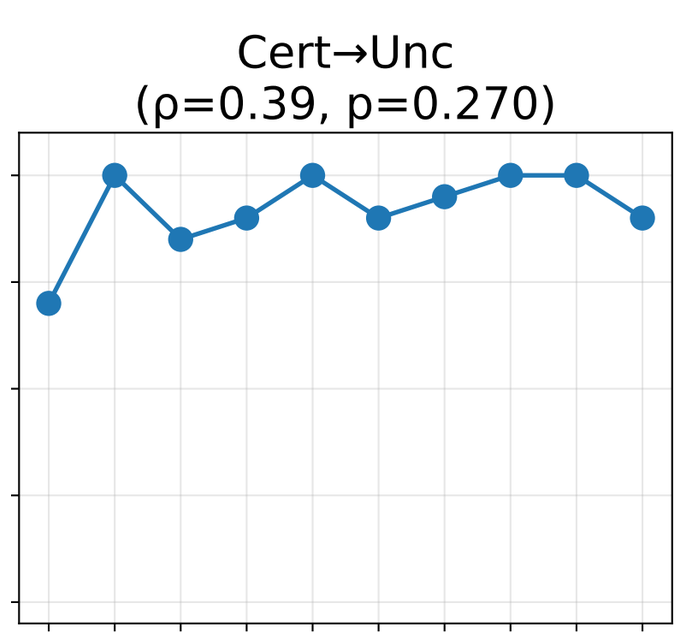
Score Trends by Release Date (Claude)

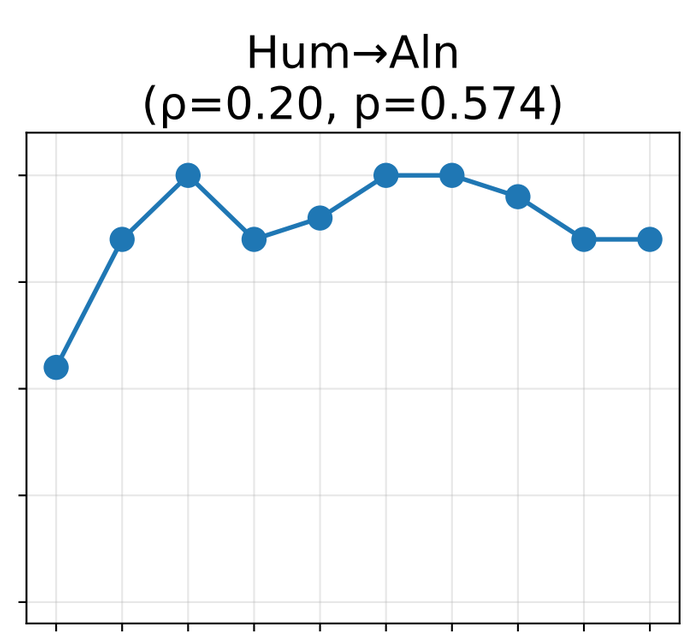




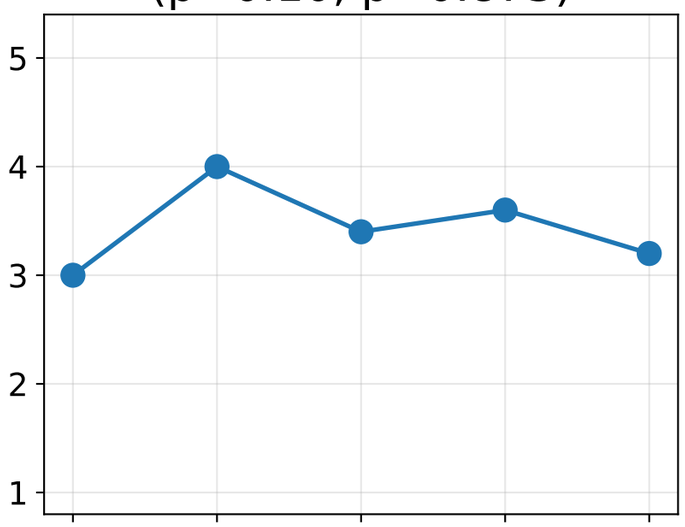


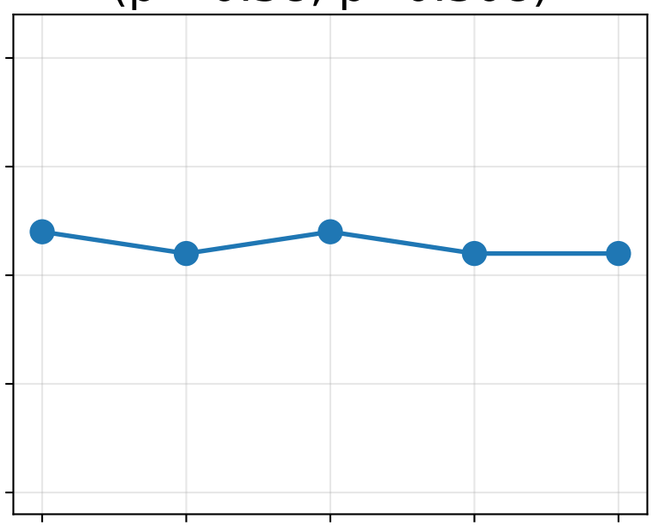


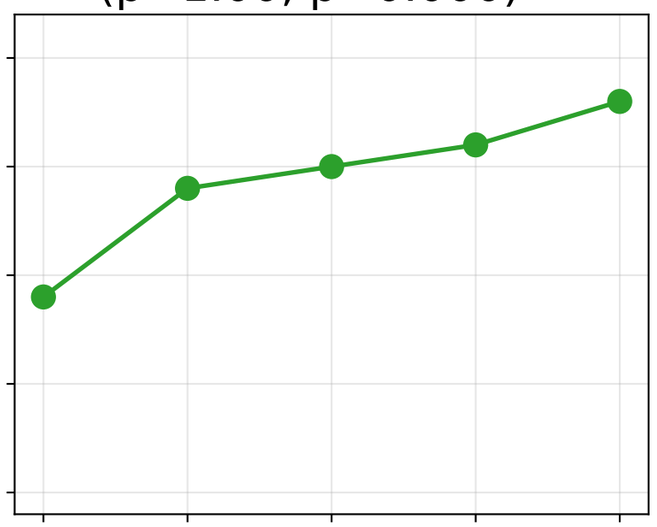


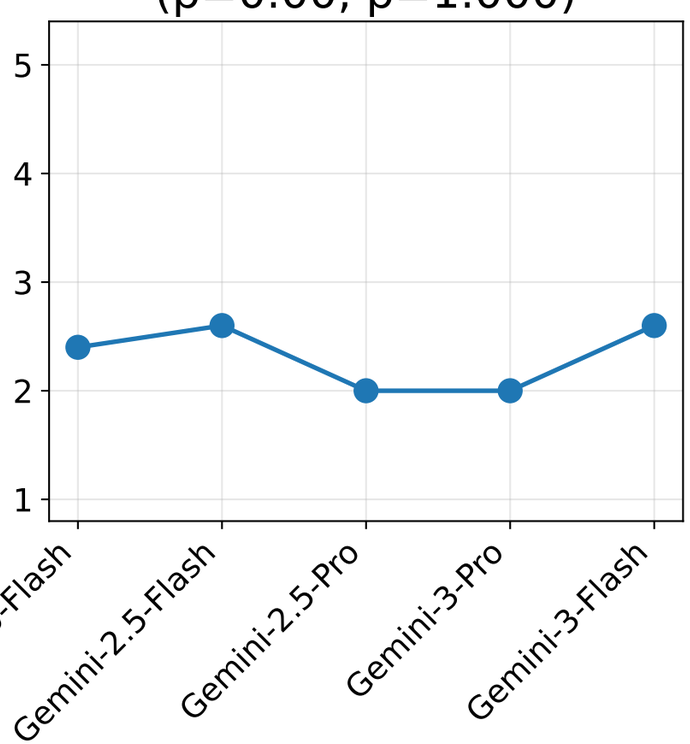


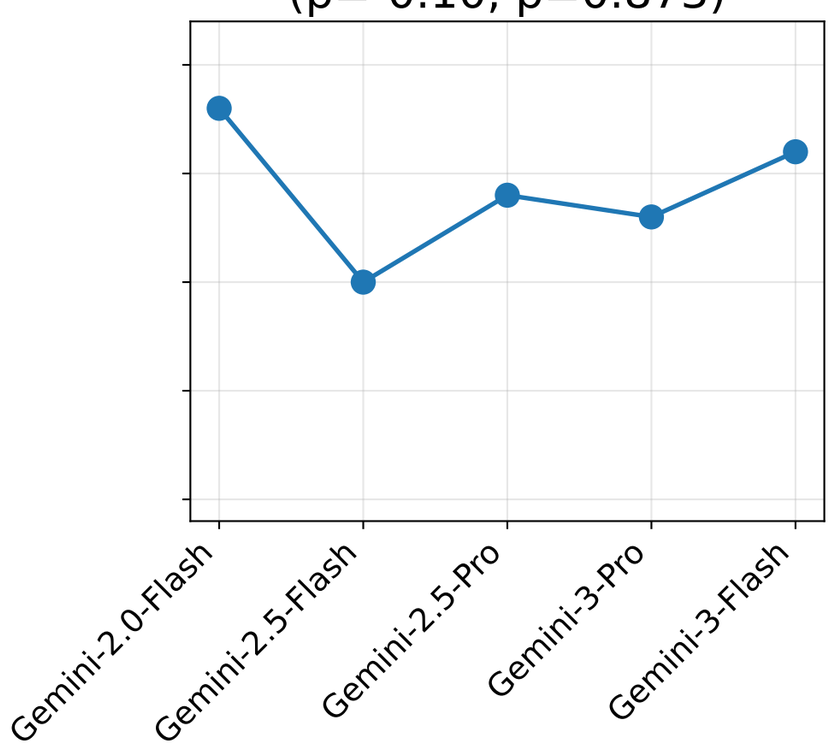
Score Trends by Release Date (Gemini)

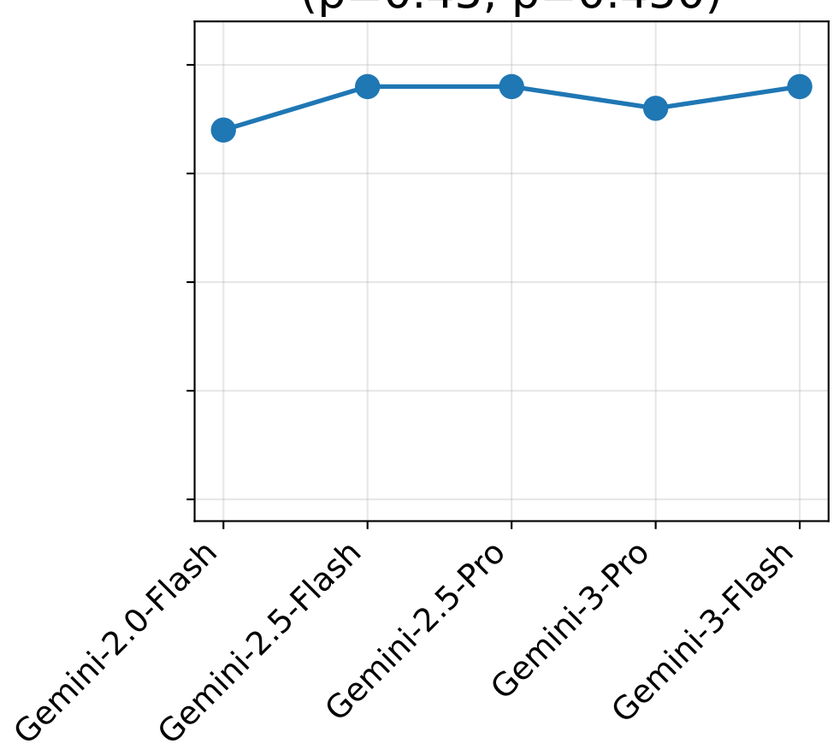




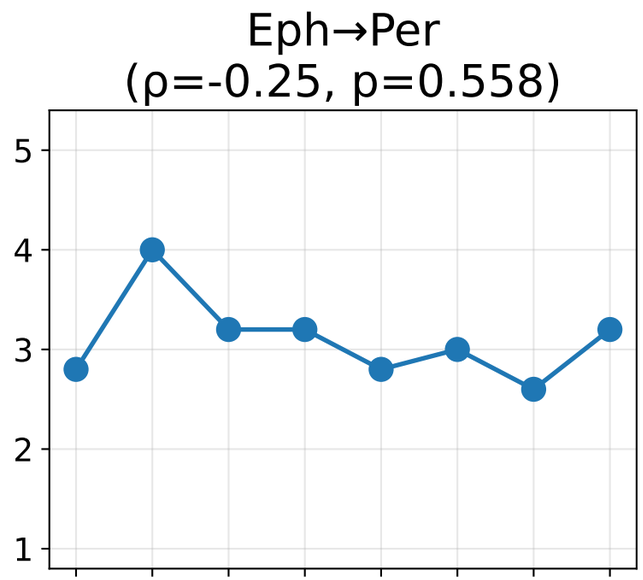


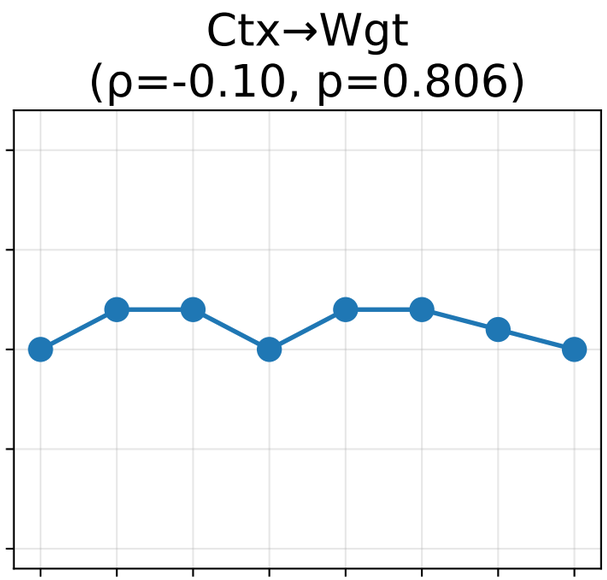


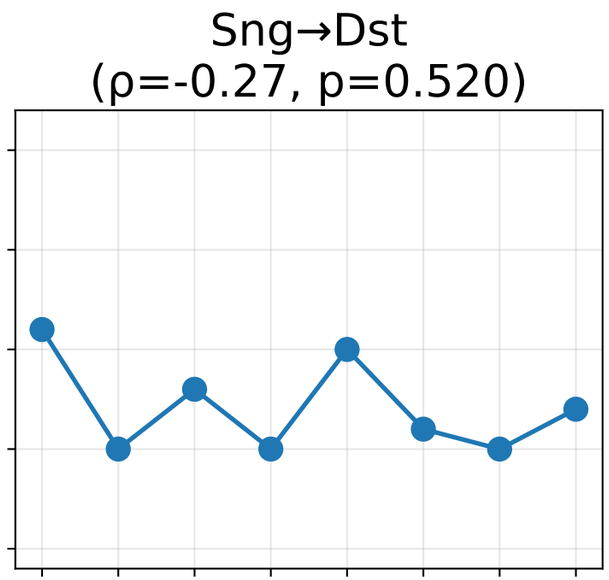


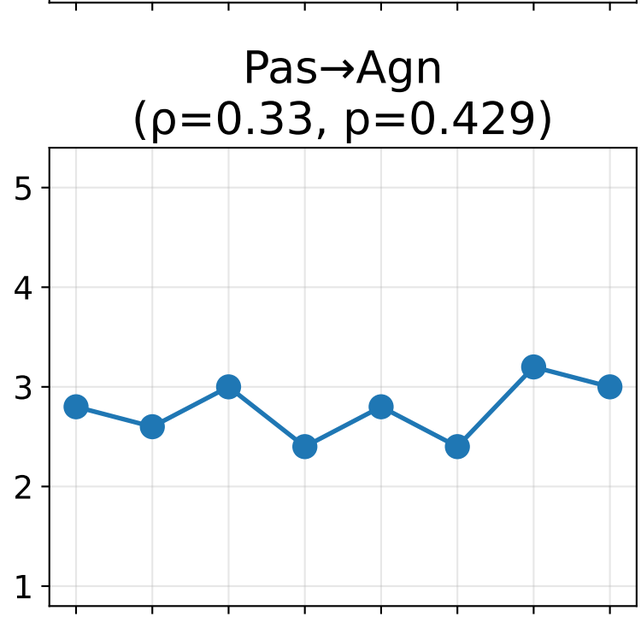


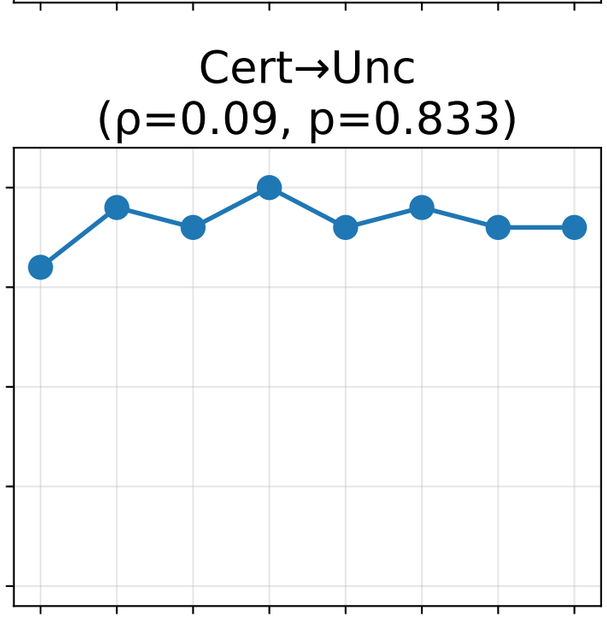
Score Trends by Release Date (Llama)

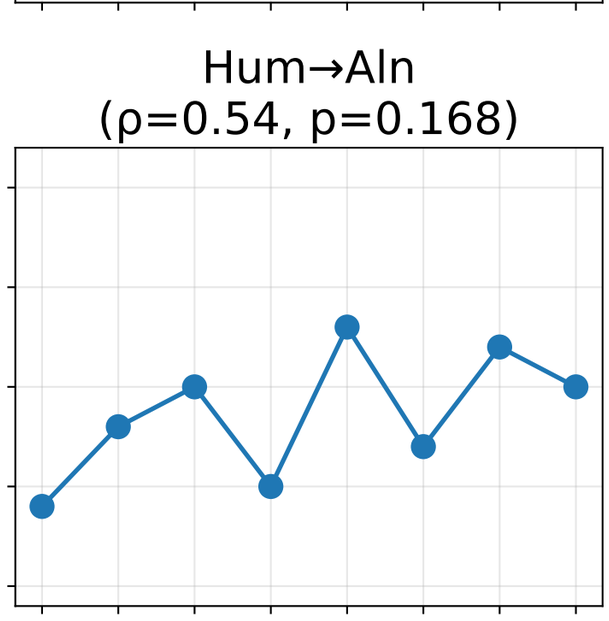












Score Trends by Release Date