Table S1. Sample characteristics (N = 14, primary site)

Characteristic	Value
N (questionnaire / eye-tracking)	14 / 13 (P08: hardware failure)
Condition	Full Display n=7, On-Demand n=7
Gender	8 women, 6 men
Age	M=39.3, SD=12.2, range 26–63
Education	7 Master's, 5 Doctorate, 1 Bachelor's, 1 Other
AR experience	1 daily, 4 often, 2 sometimes, 5 rarely, 2 never
Heritage interest (1–5)	M=3.36, SD=0.93
Prescription glasses	8 yes, 6 no

Two pilot sessions at a second site (AI-generated imagery) are excluded from all data and analyses; see paper Sec. 5.1.

Table S2. AOI dwell % by condition — exact Mann-Whitney U per task and AOI group

Task	AOI group	n_F	n_OD	Mdn_F	Mdn_OD	U	p	r_rb	95% CI (r)
T0	Static	6	6	18.94	35.80	4.0	0.026	-0.78	[-1.00, -0.33]
T0	Integrated	6	6	64.05	55.11	31.0	0.041	0.72	[0.17, 1.00]
T0	Detail	6	6	4.75	1.50	25.0	0.295	0.39	[-0.33, 0.94]
T0	Environment	6	6	6.53	9.09	16.0	0.818	-0.11	[-0.83, 0.56]
T1	Static	6	6	83.01	100.00	13.0	0.455	-0.28	[-0.83, 0.39]
T1	Integrated	6	6	0.00	0.00	20.0	0.753	0.11	[-0.42, 0.67]
T1	Detail	6	6	0.00	0.00	18.0	1.000	0.00	[0.00, 0.00]
T1	Environment	6	6	9.25	0.00	26.0	0.200	0.44	[-0.14, 1.00]
T2	Static	6	7	55.12	59.42	16.0	0.534	-0.24	[-0.81, 0.43]
T2	Integrated	6	7	31.43	19.63	29.0	0.295	0.38	[-0.29, 0.90]
T2	Detail	6	7	12.83	4.11	30.0	0.222	0.43	[-0.24, 0.95]
T2	Environment	6	7	0.00	0.00	15.0	0.337	-0.29	[-0.71, 0.21]
T3	Static	6	7	26.89	38.04	18.5	0.775	-0.12	[-0.74, 0.57]
T3	Integrated	6	7	61.05	46.63	23.0	0.836	0.10	[-0.57, 0.76]
T3	Detail	6	7	8.12	6.41	18.0	0.731	-0.14	[-0.76, 0.57]
T3	Environment	6	7	7.74	5.17	26.5	0.474	0.26	[-0.48, 0.90]
T4	Static	5	7	12.71	30.19	10.0	0.268	-0.43	[-1.00, 0.26]
T4	Integrated	5	7	54.60	34.68	21.0	0.639	0.20	[-0.54, 0.89]
T4	Detail	5	7	3.23	1.88	17.0	1.000	-0.03	[-0.66, 0.66]
T4	Environment	5	7	7.80	8.43	17.0	1.000	-0.03	[-0.74, 0.66]

r_rb = rank-biserial correlation (positive: Full Display > On-Demand); CI via seeded bootstrap (10,000 resamples).

Table S3. Between-subjects family (RQ1/RQ3), Holm-corrected

Outcome	Mdn_F	Mdn_OD	U	p	r_rb	95% CI (r)	p_Holm
Workload (RTLX, T0–T	2.40	3.87	12.0	0.128	-0.51	[-0.96, 0.06]	0.513
lutter composite (T0-	2.47	2.93	22.0	0.805	-0.10	[-0.76, 0.55]	1.000
ES overall (PU revers	4.17	3.83	32.0	0.367	0.31	[-0.33, 0.86]	1.000
Total time T1–T4 (s)	307.00	338.00	22.0	0.805	-0.10	[-0.71, 0.55]	1.000

UES overall shown for completeness; its composite reliability is poor (alpha = .13) and the paper analyzes subscales separately.

Table S4. NASA-TLX subscales (participant means over T0–T4)

Subscale	M_Full	M_OD	U	p	r_rb
Mental Demand	4.31	5.94	15.0	0.259	-0.39
Physical Demand	2.91	4.46	16.5	0.332	-0.33
Temporal Demand	3.77	5.06	20.5	0.653	-0.16
Performance	4.31	7.69	17.0	0.370	-0.31
Effort	4.89	5.91	19.5	0.564	-0.20
Frustration	2.94	4.03	13.5	0.178	-0.45

All six subscales point the same direction (Full < On-Demand); none is individually significant.

Table S5. Clutter composite by task (reverse-coded, 1–7, alpha = .85)

Task	n_F	n_OD	Mdn_F	Mdn_OD	U	p	r_rb	95% CI (r)
T0	7	7	3.33	1.67	31.0	0.435	0.27	[-0.35, 0.84]
T1	7	7	2.00	2.00	23.5	0.948	-0.04	[-0.67, 0.59]
T2	7	7	2.00	1.67	29.5	0.540	0.20	[-0.41, 0.73]
T3	7	7	2.67	4.67	11.0	0.096	-0.55	[-1.00, 0.04]
T4	7	7	2.00	2.33	18.0	0.441	-0.27	[-0.84, 0.35]

Table S6. UES-SF subscales (1–5; PU items reverse-scored)

Subscale	M	SD	U (F vs OD)	p	r_rb
Focused Attention	3.69	0.78	29.5	0.559	0.20
Perceived Usability (rev.)	3.83	0.85	24.0	1.000	-0.02
Aesthetic Appeal	4.07	0.54	24.5	1.000	0.00
Reward	4.55	0.52	29.0	0.580	0.18

Friedman across subscales: chi2(3) = 9.82, p = 0.020, W = 0.23.

Table S7. Task completion time (s) by condition

Task	n_F	n_OD	Mdn_F	Mdn_OD	U	p	r_rb	95% CI (r)
T0	7	7	200.00	238.00	11.5	0.110	-0.53	[-1.00, 0.06]
T1	7	7	19.00	37.00	16.0	0.306	-0.35	[-1.00, 0.39]
T2	7	7	39.00	29.00	33.0	0.306	0.35	[-0.31, 0.88]
T3	7	7	172.00	198.00	21.0	0.710	-0.14	[-0.76, 0.55]
T4	7	7	32.00	47.00	23.0	0.902	-0.06	[-0.71, 0.71]

Within-subjects task effect (T1–T4, n = 14): Friedman chi2(3) = 21.63, p < .001, W = 0.52. T0 is a fixed ~3-minute exploration; its 'completion time' is not meaningful.

Table S8. Interface preference rankings (n = 12)

Interface	Mean rank	Ranked 1st
Map + Timerange	1.50	7
Temporal Lens	2.08	3
Temporal Distribution (Histogram)	2.42	2

Friedman chi2(2) = 5.17, p = 0.076, W = 0.22. Two participants (P08, P09) declined to rank, arguing all three interfaces were needed in combination.