European Journal of Radiology Artificial Intelligence 6 (2026) 100076

Contents lists available at ScienceDirect

European Journal of Radiology Artificial Intelligence

journal homepage: www.sciencedirect.com/journal/european-journal-of-radiology-artificial-intelligence

Impact of AI assistance on knee MRI reading time: A real-world multicenter study

Benoit Rizk [a], Natalie Heracleous [a,*], Benoit Dufour [a], Cyril Thouly [a], Sergey Morozov [a], Philippine Cordelle [b], David Goyard-Lacroix [b], Guillaume Herpe [b], Olivier Berrebi [a], Johann Carrard [a], Daniela Duarte Moreira [a], Luca Duc [a], Maria Firsova [a], Arnaud Gregoire [a], Marc Mazilu [a], Aphrodite Syrogiannopoulou [a], Federica Zanca [c]

[a] 3R Swiss Imaging Network SA, Rue de la Dixence 8, Sion CH-1950, Switzerland
[b] Incepto Medical, 10, Boulevard de Grenelle, Paris 75015, France
[c] Palindromo Consulting, Willem de Croylaan 51, Leuven 3001, Belgium

ARTICLE INFO

Keywords:
Artificial intelligence
Radiology workflow
Knee MRI
Structured reporting
Reporting time, Workflow efficiency
Time gain

ABSTRACT

Background: Artificial intelligence (AI) has the potential to enhance radiology efficiency, but real-world data on its effect on reporting workflows is still limited.This study evaluates the impact of seamless AI integration on knee MRI reporting time in a real-world, multicenter radiology setting comparing standard reporting, partially integrated AI support, and fully AI-assisted structured reporting workflows.

Methods: We conducted a prospective, multi-center study across ten Swiss radiology centers from September 2023 to October 2024. Eight radiologists (four generalists, four MSK subspecialists) interpreted 1285 knee MRI exams over three sequential phases: (1) standard reporting without AI, (2) partial AI integration with image-only findings, and (3) full AI-assisted structured reporting with auto-populated templates. Reporting time was automatically recorded via the RIS radiology information system. Differences across phases were evaluated using the Kruskal–Wallis test and post-hoc Dunn's tests. An inverse-variance weighted approach was used to compare mean reporting time differences. A post-study survey evaluated user satisfaction.

Results: Fully integrated AI-assisted reporting (Phase 3) significantly reduced average reporting time by 13.4 % ($p < 0.01$) compared to baseline. General radiologists benefited most (17.5 % reduction, $p < 0.01$), while MSK subspecialists experienced smaller improvements. Partial AI integration (Phase 2) did not reduce reporting time and occasionally increased it. User feedback indicated that 75 % found AI helpful and 62.5 % appreciated structured reporting integration.

Conclusions: Fully integrated AI-assisted structured reporting significantly reduced knee MRI interpretation time, especially for general radiologists. These findings support broader implementation of AI in routine musculoskeletal imaging workflows.

Key Points:

- AI-integrated reporting reduces knee MRI reporting time
- General radiologists benefit most from AI assistance
- Partial AI use without workflow integration may slow reporting
- Radiologists report high satisfaction with structured AI reports
- Effective AI embedding is essential for clinical time savings

* Corresponding author.
E-mail addresses: benoit.rizk@groupe3r.ch (B. Rizk), natalie.heracleous@groupe3r.ch (N. Heracleous), benoit.dufour@groupe3r.ch (B. Dufour), cyril.thouly@groupe3r.ch (C. Thouly), sergey.morozov@groupe3r.ch (S. Morozov), philippine.cordelle@incepto-medical.com (P. Cordelle), david.goyard-lacroix@incepto-medical.com (D. Goyard-Lacroix), guillaume.herpe@incepto-medical.com (G. Herpe), olivier.berrebi@groupe3r.ch (O. Berrebi), johann.carrard@groupe3r.ch (J. Carrard), daniela.duartemoreira@groupe3r.ch (D.D. Moreira), luca.duc@groupe3r.ch (L. Duc), maria.firsova@groupe3r.ch (M. Firsova), arnaud.gregoire@groupe3r.ch (A. Gregoire), marc.mazilu@groupe3r.ch (M. Mazilu), aphrodite.syrogiannopoulou@groupe3r.ch (A. Syrogiannopoulou), federica.zanca@palindromo.consulting (F. Zanca).

https://doi.org/10.1016/j.ejrai.2026.100076
Received 13 November 2025; Received in revised form 13 January 2026; Accepted 6 February 2026
Available online 20 February 2026

1. Introduction

Artificial intelligence (AI) is gaining momentum in the field of radiology workflow efficiency and it is increasingly used to support radiologists in time-consuming diagnostic and non-diagnostic tasks [1]. Even though there are currently over 200 CE-marked AI tools available [2–4], their full potential is often limited by inadequate integration into clinical workflows. Many current solutions require radiologists to review results in separate windows and transfer findings into their reports, such as using voice recognition to integrate AI outputs, a process that increases complexity without improving efficiency [5,6].

This disconnect between AI output and radiologist workflow not only undermines the efficiency gains promised by these tools but also highlights the urgent need for more seamless and intelligent integration strategies.

Effective integration into the radiologist's daily routine is essential to unlock the true potential of AI. Poorly integrated tools can disrupt reading workflows, increase cognitive load, and elevate the risk of errors. Moreover, when AI-generated findings are stored in PACS without links to structured reports and databases, their utility for research, audit, and quality improvement remains limited.

In musculoskeletal imaging (MSK), AI algorithms such as KEROS (Incepto Medical™, Paris, France) [7,8] are designed to assist in knee MRI interpretation by automatically assessing structures such as cartilage, menisci [9] and ligaments [10]. KEROS employs deep learning models trained on a diverse dataset of knee MRIs to detect abnormalities and prepopulate structured reports with its findings [10]. This automation reduces the cognitive burden on radiologists and allows for standardized, high-quality reporting.

To address the integration gap highlighted in the introduction (where AI results are often reviewed separately and not embedded in radiology reports) we investigated a new reporting workflow in which AI-generated findings from KEROS are directly integrated into structured knee MRI reports.

We evaluated prospectively the impact of this integrated workflow on reporting time among radiologists with varying levels of experience and subspecialty training across multiple clinical centers in Switzerland. While previous studies have demonstrated the potential of AI in image analysis, few have explored its practical implementation and measurable effects on radiology workflow [11,12]. Building on this gap in the literature, this study aims to assess how AI-assisted structured reporting influences interpretation time in a real-world clinical setting.

2. Materials and methods

All subjects gave their informed consent for inclusion before they participated in the study. All procedures were complied with local regulatory and institutional standards. No personal patient data were used, as the study primarily focuses on the radiologists' reading time.

2.1. Study design

A prospective, multi-center study was conducted between September 2023 and October 2024 across ten radiology outpatient centers in Switzerland (3 R Swiss Imaging Network, Groupe 3 R). Eight radiologists participated, including four musculoskeletal (MSK) subspecialists and four general radiologists (GEN).

The radiologist group included five males and three females. Most radiologists have between 6 and 15 years of experience (75 %), while the remaining 25 % have 3–5 years of experience. The monthly reading volume of knee MRIs per radiologist ranges from 30 to 54 exams, with musculoskeletal specialists generally interpreting higher volumes.

The study included patients selected based on predefined inclusion criteria: older than 16 years, who underwent knee MRI scans for diverse indications across all severity levels. Cases were excluded if relevant clinical data were missing, if there were motion artifacts detected or lack of T2w sequences. The quality and effectiveness of KEROS were not analyzed in this study, as they have been previously demonstrated in peer-reviewed scientific publications [7–10]; we did not assess its performance or conduct subgroup analyses. Cases outside KEROS intended use, such as post-operative knee exams where AI does not normally deliver an output, were excluded from the study. Given the extensive diversity of pathologies observed within our network (as detailed in the results), the findings are considered to be generalizable across a wide range of healthcare settings.

Data was collected consecutively over three distinct phases: Phase 1 - Baseline: Standard reporting without any AI support. Phase 2 - Partial AI Integration: Radiologists had access to AI findings via a separate interface but needed to transcribe or interpret results with speech-recognition device. Phase 3 - Full AI Integration: AI-generated findings were automatically integrated into structured reporting templates, enabling direct validation and editing by radiologists. The included exams were different in each phase. Fig. 1 illustrates the flowchart of the study consisting in the three different phases of the study along with the three AI implementation workflows.

We conducted the study in real clinical settings, where case allocation naturally follows individual workload rather than controlled randomization. As a consequence, each radiologist reviewed a set of cases in all three phases, allocated according to their clinical workload. There was no specific scheduling to ensure a balanced distribution of case complexity and modality between the radiologists. Reporting time was the primary outcome measure and was captured automatically using time-tracking features in the reporting system (RIS Xplore 7.2.32). Secondary outcomes, such as user satisfaction and overall study efficiency, were assessed through a post-study questionnaire. KEROS's effectiveness was already confirmed through CE certification processes that validated analytical and clinical performance. Our institutional quality management system, supported by peer-review and peer-learning frameworks, ensures ongoing quality assurance without requiring reference standard validation for this workflow efficiency study. This methodological approach is consistent with regulatory guidance distinguishing pre-market diagnostic validation from post-market operational assessment, and follows established practices for evaluating CE-certified AI medical devices in clinical workflows.

Radiologists were not blinded to the study objectives, as they were requested not to be interrupted during their exam reading to minimize performance and observation bias. A flag was available to indicate if any interruptions occurred, and flagged cases were excluded from the analysis. Radiologists were also instructed not to open the images prior to reporting, as explicitly explained through videos, documentations, and live demonstration. No feedback on timing or performance was provided during data collection. An a priori power analysis for a one-way ANOVA indicated that a minimum of 352 cases per phase was required to achieve statistical significance.

To improve the quality of the preselected data, a filter was applied to exclude outliers with reading time longer than 50 min, such as cases affected by technical issues or interruptions. This threshold was established in consultation with radiologists as the maximum acceptable duration for uninterrupted readings

Phase 1 (04/09/2023–28/11/2023 and 27/08/2024–07/10/2024)

In Phase 1, radiologists reported knee MRI examinations using voice-recognition dictation to fill standardized structured templates. The workflow included the radiologist opening the MRI examination via the RIS system and then reviewing the images using the PACS. Findings were dictated in the radiology report through the RIS system (RIS Xplore 7.2.32) using the integrated voice recognition software (Dragon Medical One - DMO) which is currently used in clinical routine. Final edits and formatting were performed manually by the radiologists. The interface used in this phase is a standard reporting system with no AI integration as shown in Fig. 2a. This phase serves as the baseline, where radiologists only rely on their expertise and experience. The sample size included 431 exams, chosen to be above the minimum number of cases indicated

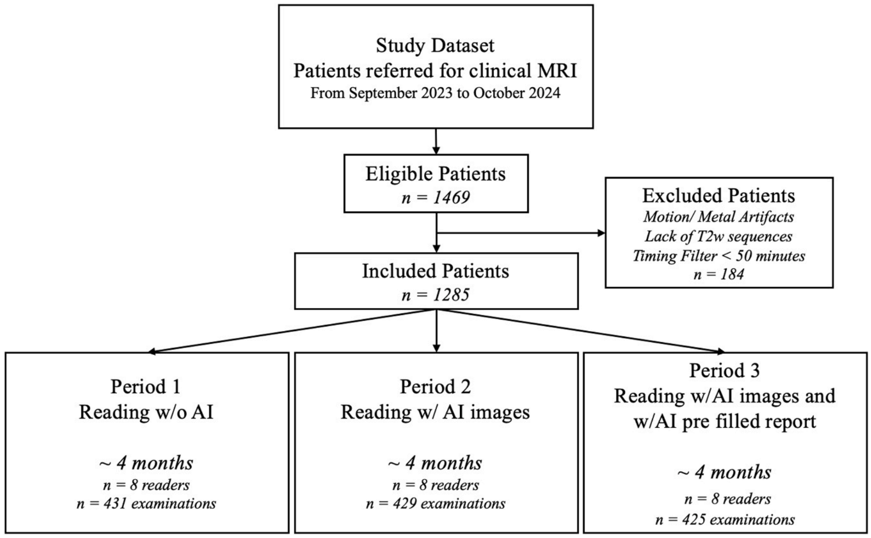


Fig. 1. Flowchart which summarizes the three phase implementation workflows (w/o without, w/ with).

by the power analysis conducted prior to the study.

Phase 1, was divided in two parts; the first one was at the beginning of this study while the second one was after Phase 3 was concluded. The reason for that was twofold; to compensate for the experience gained in time by radiologists but also to have similar number of cases as the other two phases, as indicated by the power analysis. The time periods are given in detail in Table 1.

Phase 2 (29/11/2023–14/04/2024)

In Phase 2, radiologists have access to AI-generated diagnostic suggestions before initiating the report. In the same way as in Phase 1, radiologists accessed the patient's MRI examination via the PACS and reported using the RIS using voice recognition software. The KEROS AI-generated diagnostic findings were displayed as an additional tab within the PACS interface, allowing radiologists to review AI predictions alongside their interpretation but without any automated integration into the report, as can be shown in Figs. 2b and 2c. These suggestions are displayed separately from the report and are not integrated into the reporting system. Radiologists were free to choose when to consult the AI results, whether before reviewing the MRI images or afterward. Radiologists manually edited and formatted the final report. This phase is designed to allow radiologists to evaluate the AI's potential with minimum changes from their typical workflow. The sample included 429 exams. Timer began when the report was opened in the RIS and ended upon final report validation.

Phase 3 (15/04/2024–25/08/2024)

Finally, in Phase 3, radiologists accessed the MRI examination via the PACS and reporting via the RIS, as in the previous phases. However, in this phase KEROS AI-generated findings were integrated directly into the structured radiology report (SSR) by automatically pre-filling the structured report (SSR) template in the RIS. The template used 18 predefined sections, each corresponding to a specific anatomical region or finding. Report data were transmitted using HL7 ORU messages- a standard format for medical results- in which each OBX (Observation/Result) segment carried a distinct AI-generated observation. All 18 sections of the report were pre-filled with normal findings and 15 out of 18 could be automatically completed without radiologist input. An iterative process was used to fine-tune the layout, wording, and punctuation of the SSR, with the goal of assessing how many fields could be reliably auto-filled.

Radiologists were able to review, edit, and validate the AI pre-filled content before finalizing the report. Edits could be made via voice recognition or direct typing, with minimal manual formatting required, allowing radiologists to use their preferred method. Fig. 2d illustrates layout of the imaging and reporting tools during this phase. Phase 3 represents a fully integrated AI reporting workflow, with a sample of 425 MRI exams. In this phase, confirming AI findings or overwriting pre-filled sections introduced one additional interaction step compared to the previous phases. Illustrative examples of cases where AI significantly helped both the quality and time to routine reporting can be found in Annex 2.

2.2. Statistical analysis

For the calculation of the mean reporting time, $\overline{RT}^{RAD}_{Phase}$, for a given radiologist in a specific phase, the following formula is used:

$$\overline{RT}^{RAD}_{Phase} = \frac{1}{n^{RAD}_{Phase}} \sum_{i=1}^{n^{RAD}_{Phase}} RT^{RAD}_{iPhase}$$

where RAD indicates the radiologist (1–8), $Phase$ is one of $P1, P2, P3$ indicating the phase of the study (1,2 or 3, respectively), RT^{RAD}_{iPhase} is the reporting time for case i read by that radiologist in that phase, and n^{RAD}_{Phase} is the number of exams analyzed by a given radiologist for a given phase.

To estimate the average difference in reporting time between phases, we applied the inverse-variance weighting method [13]. In this approach, each radiologist's contribution to the group estimate is weighted inversely to the variance of their individual phase difference, giving greater influence to more precise (i.e., lower-variance) estimates. The weighted average difference between two phases, for example between Phase 2 and Phase 1, is calculated as follows:

$$\widehat{\Delta}_{12} = \frac{\sum_{RAD=1}^{8} \left(\Delta^{RAD}_{12} \middle/ \sigma^{2\,RAD}_{12} \right)}{\sum_{RAD=1}^{8} \left(1 \middle/ \sigma^{2\,RAD}_{12} \right)}$$

where,

$$\Delta^{RAD}_{12} = \overline{RT}^{RAD}_{P2} - \overline{RT}^{RAD}_{P1}$$

is the mean difference in reporting time for each radiologist between Phase 2 and Phase 1, and

$$\sigma^{2\,RAD}_{12} = VAR\left(\overline{RT}^{RAD}_{P2}\right) + VAR\left(\overline{RT}^{RAD}_{P1}\right)$$

represents the combined variance of the phase-specific means for each radiologist.

If variances are assumed to be equal across all radiologists, this formula simplifies to an unweighted arithmetic mean of the differences.

Fig. 2. (a) Example of the Phase 1 interface: radiology report structure (left) and MRI knee images (right). (b) Phase 2 interface: AI findings displayed on the right, with no integration into the structured report template on the left. (c) PACS environment in Phase 2: AI PDF report and schema/DICOM secondary capture shown with color-coded hanging protocols. (d) Phase 3 RIS/PACS interface: AI output integrated into the structured report before radiologist interaction/editing via voice input.

Table 1
Total number of exams conducted during three distinct phases, along with the biological sex distribution (count and percentage) of male and female participants for each phase.

Phase	Time Period	Total number of exams	Male (Count, %)	Female (Count, %)
Phase 1	04/09/2023–28/11/2023 27/08/2024–07/10/2024	431	194 (45.1 %)	237 (54.9 %)
Phase 2	29/11/2023–14/04/2024	429	198 (46.1 %)	231 (53.9 %)
Phase 3	15/04/2024–25/08/2024	425	213 (50.1 %)	212 (49.9 %)

However, the inverse-variance approach allows for more accurate estimation when the precision of the radiologists' reporting time measurements varies.

This approach was chosen based on several factors. First, it was important to recognize that each radiologist handled a different number of cases, which could introduce variability in their individual contributions to the overall results. Moreover, the same radiologist may have a different number of cases across the three phases, adding another layer of complexity when determining how to apply weights. In addition, this method helps in avoiding single outliers in each phase dominating the result. Therefore, to protect against biasing the results, for example, if a radiologist takes very long to read their exams. Consequently, we treated each individual radiologist as an independent experiment, rather than using each individual measurement.

To assess whether the differences in reporting times across the three reporting phases were statistically significant, we applied the Kruskal-Wallis H test, a non-parametric alternative to one-way ANOVA. This test was chosen due to the non-normal distribution of reporting times (see results). The Kruskal-Wallis test was used to compare the distribution of reporting times across Phase 1, Phase 2, and Phase 3 for the overall dataset, as well as within each radiologist expertise category (GEN and MSK). Post-hoc Dunn's test for pairwise comparisons with Holm corrections was also conducted. Descriptive statistics including patient sex distribution per phase and radiologist expertise were also calculated.

2.3. Qualitative evaluation

Following study completion, a qualitative evaluation was carried out through a structured questionnaire, delivered via Google Forms, to collect feedback from all eight participating radiologists. The questionnaire assessed key aspects, such as workflow integration, usability of the KEROS software, performance of the AI-assisted reporting process, and the overall effectiveness across the three phases of the study. Results from this evaluation are presented in the Results section, and the full list of the questions is provided in the Annex 1.

3. Results

3.1. Data collection

A total of 1285 knee MRI exams were included across three phases (Table 1). Statistically significant differences in distributions across phases were observed, with results indicating variation among male ($p < 0.001$) and female ($p < 0.001$) participants. In a similar way, as shown in the data in Table 2, statistically significant differences in reporting times across phases were observed for all radiologists combined ($p < 0.001$), as well as within subgroups: GENs ($p < 0.001$) and MSKs ($p < 0.001$).

The distribution of the datasets can be shown in Fig. 3. The Shapiro-Wilk test was conducted for testing the normality within each phase

Table 2
Total number of exams and the distribution between general radiologists (GEN) and musculoskeletal radiologists (MSK) across three study phases. The average reporting time $\overline{RT}_{Phase}^{RAD}$ in minutes and corresponding standard deviations for each group.

Phase	Experience Level	Number of Exams, %	Average reporting time ($\overline{RT}_{Phase}^{RAD}$) [min]	Standard deviation [min]
Phase 1	Total	431	17.0	9.81
	GEN	189 (44 %)	17.7	11.4
	MSK	242 (56 %)	16.4	8.37
Phase 2	Total	429	19.5	10.1
	GEN	193 (45 %)	19.6	11.3
	MSK	236 (55 %)	17.7	9.41
Phase 3	Total	425	14.0	7.36
	GEN	174 (41 %)	14.2	8.54
	MSK	251 (59 %)	13.9	6.42

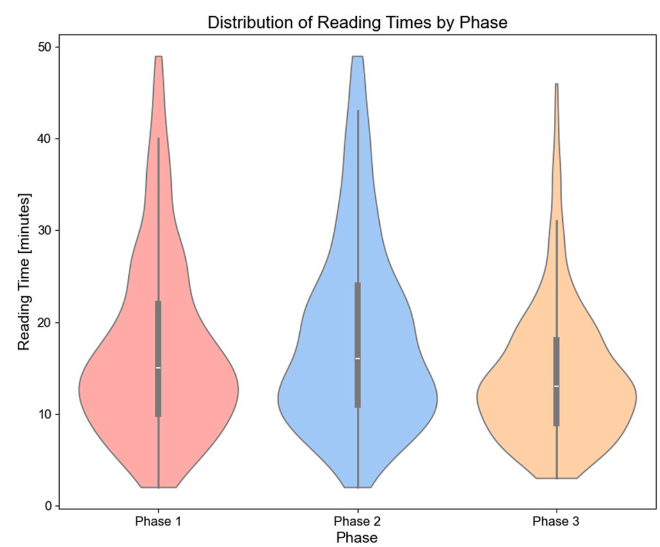


Fig. 3. Violin plots illustrating the distribution of reading times (in minutes) across three phases. Each violin includes a mirrored histogram with an embedded box plot. The three distributions are not normal but with a rather long tail to right (longer timings).

group, revealing non-normal distributions. Statistically significant differences in reporting times across phases were observed. The post-hoc test results indicated statistically significant differences across all pairwise group comparisons. Specifically, delay times differed significantly between Phase 1 and Phase 2 ($p = 0.02$), between Phase 1 and Phase 3 ($p < 0.001$), and between Phase 2 and Phase 3 ($p < 0.001$). These findings suggested that indeed each group demonstrated distinct delay profiles, with the most pronounced difference observed between Phases 2 and 3.

No adverse events or diagnostic concerns were reported through institutional quality assurance systems during the study period.

3.2. Reporting time for each phase

The average reporting time per radiologist $\overline{RT}_{Phase}^{RAD}$ varied across the three phases. In Phase 1(standard reporting), time ranged from 10.8 to 26.3 min. In Phase 2, where AI-assisted findings were available prior to interpretation, the range increased slightly to 11.7–27.5 min. In Phase 3, where AI-generated reports were prepopulated, average reporting times decreased to between 9.2 and 21.6 min.

Fig. 4 presents box plots of reporting times per radiologist, distinguishing between generalists (GEN) and musculoskeletal specialists (MSK). The plots highlight variations in both median reporting times and variability among radiologists. Differences in the presence and distribution of outliers further suggest individual differences in reading speed and occasional deviations from typical performance.

3.3. *Reading time differences between phases*

The average reading time difference for each of the radiologists between phases 3, 2 and 1 are shown in Fig. 5. All radiologists except one are saving time during Phase 3, so the prefilling of the AI results in the structure report is a feature that can result in a net gain of radiologist time.

The results comparing reporting ties across phases are presented in Table 3. They indicate a significant reduction in knee MRI reporting time with the use of AI-assisted structured reporting. Overall, reporting time decreased by 13.4 % (p-value <0.01), with the greatest impact observed among general radiologists, who achieved a 17.5 % reduction (p-value <0.01). Musculoskeletal specialists also demonstrated improved efficiency, though the reduction in their reporting time was smaller, likely reflecting their already optimized and specialized workflows.

Radiology reports were analyzed using the Mistral-7B large language model (LLM) to identify predefined knee pathologies. A radiologist manually labeled 15 cases to incorporate domain expertise into prompting and to refine the accuracy of abnormality definitions. Among 1285 knee MRI examinations (corresponding to 1251 unique patients), meniscal tears were observed in 794 examinations (61.8 %), chondropathy in 820 (63.8 %), bone contusions or fractures in 264 (20.5 %), LLI sprains in 201 (15.6 %), and LCA ruptures in 147 (11.4 %). No abnormalities were reported in 102 examinations (7.9 %). As multiple pathologies could be present within a single examination, categories were not mutually exclusive. Pathology distribution across the three study phases showed overall comparable patterns, with similar prevalence ranges for meniscal tears (approximately 59–65 %), chondropathy (61–66 %), LLI sprains (11–20 %), and examinations without abnormalities (6–9 %). Phase 2 demonstrated a slightly higher prevalence of LCA ruptures (∼18 %) and bone contusions/fractures (∼26 %), while the distribution of other pathologies remained consistent across the three phases. For comparison, the text analysis was also performed using a conventional NLP pipeline based on the spaCy library and the French language model fr_core_news_sm [14]. However, the LLM-based approach yielded more accurate pathology identification, and the results reported here therefore rely on the Mistral-7B model [15].

3.4. *Qualitative evaluation*

For the qualitative evaluation, participating radiologists reported their typical weekly workload: 50 % interpreted between 0 and 10 knee MRI examinations, while the other 50 % handled between 10 and 30. Regarding the perceived impact of KEROS on clinical practice, 75 % of the participants reported that its availability facilitated their work. Overall, radiologists noted that KEROS contributed positively to workflow efficiency and time management. In Phase 2, where the AI solution was integrated, the main limitation was time constraints. In contrast, Phase 3 revealed technical issues such as inconsistent report formatting and variable font usage in the conclusion sections of standardized reports.

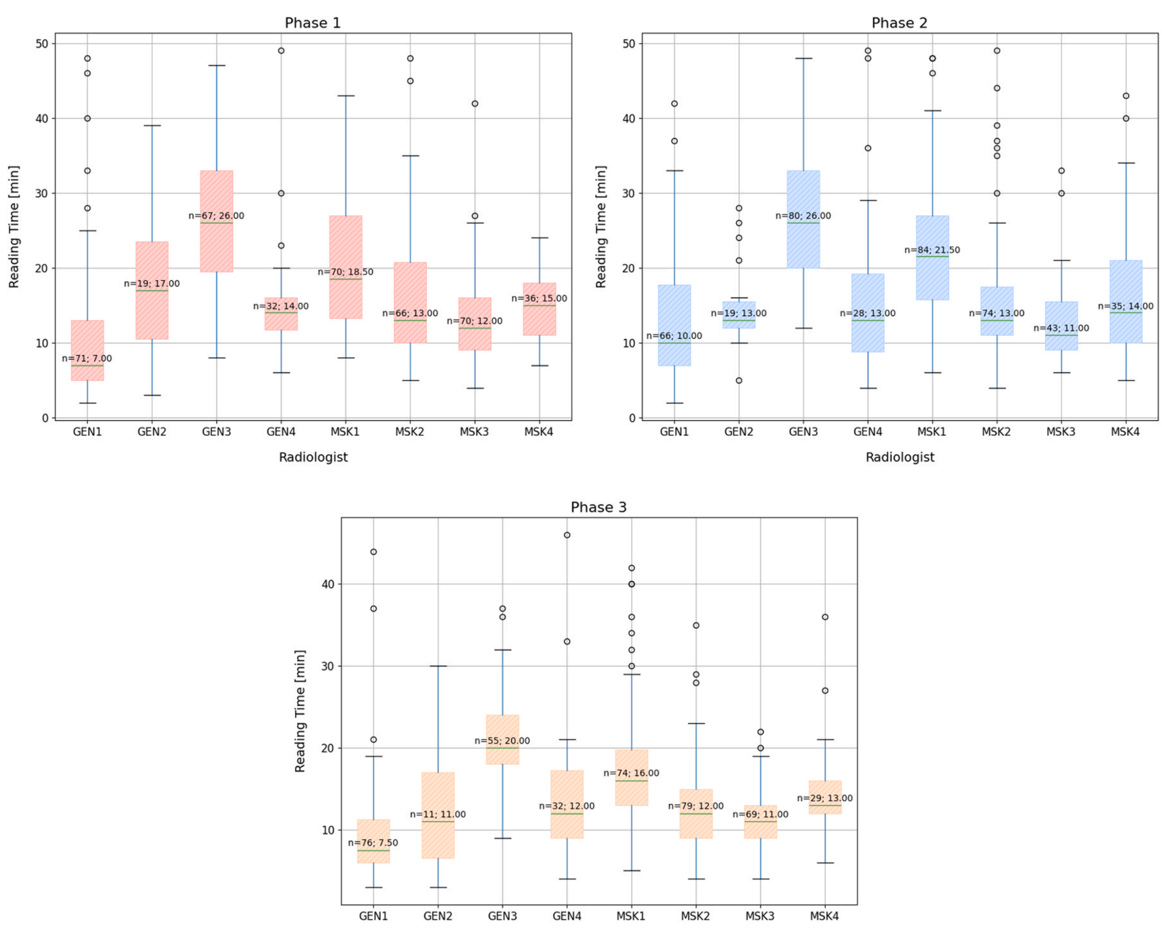


Fig. 4. Box plot representations of the reading time (RT) for each radiologist in each Phase.

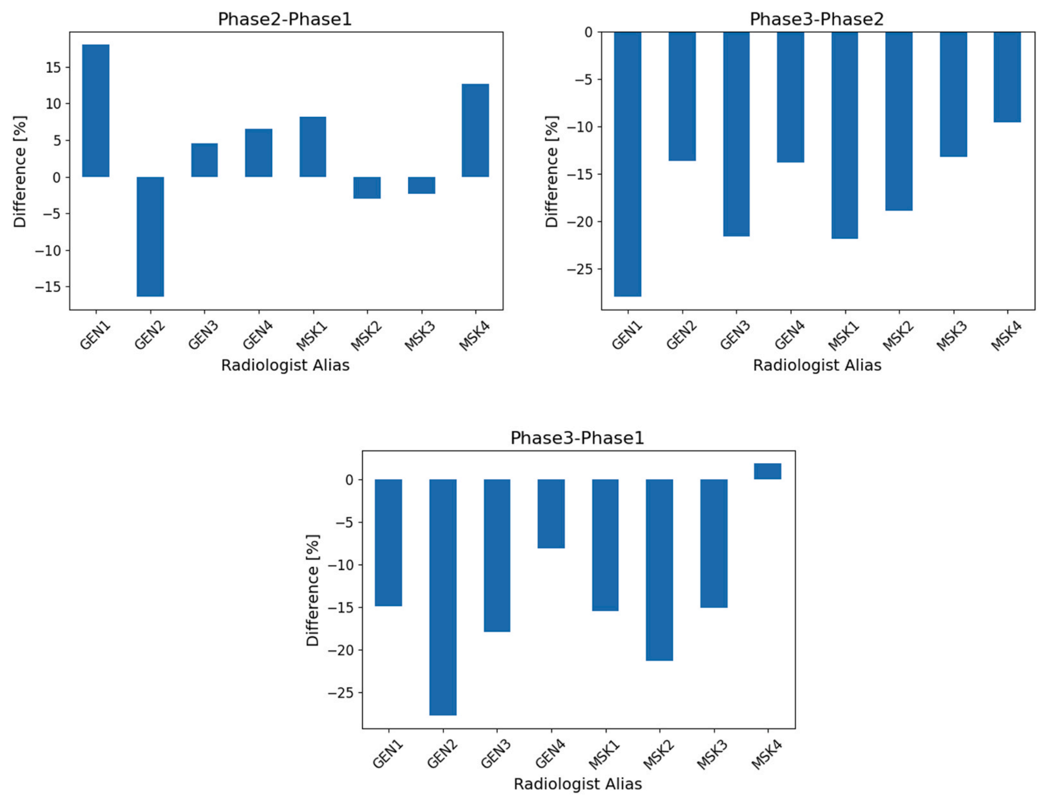


Fig. 5. Relative reporting time difference between the three phases, in percent, for each radiologist.

Table 3
The final results that show the estimated differences in time between phases. The timings are shown in minutes and the minus sign means that a net time gain was achieved.

Phase Difference	Group	Weighted Average Difference $\widehat{\Delta}$ [min]	Weighted Percentage Change [%]
Phase 2 – Phase 1	All	0.421	2.66
$\widehat{\Delta}_{12}$	GEN	0.165	0.951
	MSK	0.594	3.93
Phase 3 – Phase 2	All	-2.89	-17.7
$\widehat{\Delta}_{23}$	GEN	-3.57	-21.1
	MSK	-2.41	-15.3
Phase 3 – Phase 1	All	-2.13	-13.4
$\widehat{\Delta}_{13}$	GEN	-3.03	-17.5
	MSK	-1.62	-10.7

In terms of utility, 62.5 % of the radiologists found the integration of AI results into structured reports to be beneficial and an equal portion reported a reduction in report completion time due to AI-generated pre-filling. Suggestions for further improvement reducing false negatives, automating the insertion of key diagnostic images, and enhancing the granularity of lesion characterization- particularly for cartilage abnormalities and meniscal tears.

Opinions on the usefulness of pre-filled reports in Phase 3 varied among participants. While they were found effective for straightforward knee cases involving one or two lesions, radiologists found them less suitable for cases with diagnostic uncertainty, where frequent revisions of the AI-generated content were required. Radiologists primarily reported time savings during the pre-filling phase for simpler cases. In more complex cases, image interpretation contributed more substantially to time gains by enabling faster detection or confirmation of key findings (for example, in ski-related injuries time gain is more limited). For general radiologists, image interpretation contributed more to time savings, along with increased reassurance from AI-supported confirmation of findings.

For future development, radiologists recommended improving the characterization of structural lesions such as fissures, enhancing the prediction of joint instability, providing more precise lesion descriptions, and including illustrative key images. When asked to rate the likelihood of recommending the KEROS system to others on a scale from 1 to 10, 50 % of participants gave it a score of 8, while the remaining responses included scores of 6, 7, and 9. Notably, no scores below 6 were reported. When asked to rate the likelihood of recommending the KEROS system to others on a scale from 1 to 10, 50 % of participants gave it a score of 8, while the remaining responses included scores of 6, 7, and 9. Notably, no scores below 6 were reported.

4. Discussion

The objective of this study was to evaluate the impact of AI-assisted structured reporting on knee MRI interpretation time among radiologists with varying levels of experience and subspecialty training. Specifically, the study aimed to assess how different levels of AI integration into the reporting workflow influenced radiologists' reporting time efficiency in a real-world clinical setting across multiple centers.

The AI tool selected for the study was CE-certified and demonstrated high quality and effectiveness as shown in several published studies [7–10]. Therefore, we did not analyze the accuracy of the AI outcomes, nor did we perform subgroup analyses to evaluate which findings were missed or overcalled by the AI interpretation, and to assess the impact of false-negatives or false-positives on radiology reporting efficiency. This approach aligns with current best practices in evaluating certified AI medical devices, where workflow optimization studies assess implementation success rather than re-examining diagnostic accuracy [16–20].

The findings of this study highlight the potential of AI-assisted structured reporting to enhance radiology workflow efficiency. The most pronounced reduction in reporting time was observed among general radiologists, likely reflecting their comparatively lower baseline experience in knee MRI interpretation relative to musculoskeletal specialists. The result aligns with previous studies suggesting that AI can serve as a valuable support tool for non-specialists helping to streamline interpretation and provide structured diagnostic guidance.

The slight increase in reading time observed in Phase 2 suggests that reviewing AI-generated findings before voice reporting may introduce additional steps into the workflow. This may be attributable to integration challenges, such as the mental effort required for adaptation or misalignments with the existing workflow, as documented in other studies [21,22]. These findings emphasize that the success of AI tools in clinical practice depends not only on their capabilities but also on their efficient integration into existing workflows with minimal user interaction.

However, when AI-generated findings were directly integrated into structured reports (Phase 3), radiologists were able to significantly save time by validating and completing pre-filled information rather than dictating full reports. These findings support the implementation of AI-assisted structured reporting as an effective strategy to optimize radiology workflows. Additional examples highlighting how AI streamlined routine reporting and contributed to overall reporting time and quality are provided in Annex 2.

The change in routine for some radiologists, as well as their openness to altering daily habits, can be challenging. While AI software can offer significant support, it may not be equally beneficial to everyone, as some radiologists may not embrace new tools or changes to their established workflows.

This study's focus on workflow efficiency metrics rather than diagnostic accuracy validation, is methodologically sound and clinically appropriate for several reasons. First, KEROS underwent comprehensive pre-market validation demonstrating sensitivity ranges of 85–95 % for meniscal tears and 90–98 % for ACL injuries [7–10], establishing its diagnostic reliability prior to clinical implementation. Second, our institutional quality assurance framework, including peer-review processes and multidisciplinary consultations, provides ongoing oversight of diagnostic quality without requiring study-specific validation.

Several limitations must be considered when interpreting these findings. While diagnostic accuracy was not directly measured in this study, this approach is consistent with regulatory frameworks for evaluating CE-certified AI devices and established quality assurance practices for workflow optimization studies. Our institutional peer-review system and quality management framework provide ongoing diagnostic oversight, and no quality concerns were identified during the study period.

Real-world conditions vary across centers, including time pressure and local factors that could influence reporting times. The study sample included eight radiologists, which, while appropriate for workflow assessment, limits generalizability regarding inter-reader variability. The sequential phase design may have introduced learning effects, though the extended timeline and case diversity help mitigate this concern. We acknowledge that in our centers the median interpretation times of the radiologists are relatively long. Factors that define the individual performance and personal reporting style of the radiologists were not evaluated. Finally, a further limitation of this study is that radiologists were aware of the study objectives and the workflow phase, which may have influenced their reporting behavior and introduced potential observer bias.

Although each study phase included different cases, pathology distributions were comparable, suggesting a similar level of case complexity and limiting potential confounding in cross-phase comparisons. We are currently working on exploring the adaptive use of pre-filled reports based on case complexity, including the integration of alerts into the RIS to triage cases as complex or simple, so that they can be directed to MSK or GEN, respectively. Additionally, future research should also explore long-term adoption patterns, user satisfaction trends, and integration with different PACS/RIS systems.

CRediT authorship contribution statement

Aphrodite Syrogiannopoulou: Validation. **Sergey Morozov:** Writing – review & editing, Validation, Supervision, Resources, Investigation, Formal analysis, Data curation, Conceptualization. **Marc Mazilu:** Validation. **Cyril Thouly:** Resources, Project administration, Methodology, Funding acquisition, Conceptualization. **Arnaud Gregoire:** Validation. **Benoit Dufour:** Visualization, Software, Resources, Project administration, Data curation. **Maria Firsova:** Validation. **Natalie Heracleous:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Luca Duc:** Validation. **Benoit Rizk:** Writing – review & editing, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Daniela Moreira:** Validation. **Johann Carrard:** Validation. **Olivier Berrebi:** Validation. **Guillaume Herpe:** Project administration. **David Goyard-Lacroix:** Resources, Methodology, Investigation, Data curation. **Federica Zanca:** Writing – review & editing, Supervision, Software, Project administration, Methodology, Investigation, Data curation, Conceptualization. **Philippine Cordelle:** Methodology, Investigation, Conceptualization.

Ethical statement

Institutional Review Board approval was not required because no personal patient data was used for this analysis, only demographics and radiology report reading timings. A written informed consent was required by every patient upon admission in Groupe 3 R (a Swiss radiology services network, comprising 20 imaging centres) stating, among others, possible use of anonymized patient data for research purposes. The patient was free to oppose this use and listed as such.

Informed consent

Written informed consent was obtained from all subjects (patients) in this study.

Funding

This research was partially funded by Guerbet for R&D purposes.

Declaration of Competing Interest

The authors of this manuscript declare relationships with the following companies: Incepto-Medical. Some of the authors (PC, DG and GH) are employees of Incepto-Medical.

From the authors that are employed by 3 R Swiss Imaging Network SA, CT, BD and BR are shareholders of the Swiss Med holding.

Annex 1. Questionnaire

This annex presents the full set of questions inlcuded in the questionnaire for the participating radiologists as referenced in the paper, provided in their original published language.

1. **Mon activité** (plusieurs choix possibles)

- Radiologue généraliste
- Radiologue ostéoarticulaire

2. Par semaine, je lis:

- 0 à 10 IRMs de genou
- 10 à 30 IRMs de genou
- Plus de 30 IRMs de genou

3. **La disponibilité du KEROS a-t-elle facilité votre travail?**

- Pas du tout facilité ………. Vraiment facilité *[échelle 1–4]*

4. Selon moi, KEROS…

- m'aide à être plus sûr dans mon diagnostic *(j'ai moins souvent recours à un second avis)*
- réduit ma charge mentale *(se reconcentrer après une interruption, aller à l'essentiel, etc.)*
- facilite l'identification des structures normales
- facilite l'identification des structures pathologiques
- me permet de passer moins de temps sur la rédaction du compte-rendu

5. Quel était le "pain-point" principal pour vous lors de la phase 2? Lors de la phase 3?
6. Mon appréciation des performances de KEROS:

- LCA
- LCM
- Cartilage FTM
- Cartilage FTL
- Cartilage FP
- Ménisque médial
- Ménisque médial *(migration)*
- Ménisque latéral
- Ménisque latéral *(migration)*
- Œdème osseux
- Épanchement articulaire
- Kyste poplité
- Hauteur patellaire
- Dysplasie de trochlée
- TA-GT

7. Choisissez les 3 fonctionnalités les plus utiles selon vous:

- LCA
- LCM
- Cartilage FTM
- Cartilage FTL
- Cartilage FP
- Ménisque médial
- Ménisque médial *(migration)*
- Ménisque latéral
- Ménisque latéral *(migration)*
- Œdème osseux
- Épanchement articulaire
- Kyste poplité
- Hauteur patellaire
- Dysplasie de trochlée
- TA-GT

8. Commentaires sur les performances de KEROS *(cas particuliers, cas fréquemment ratés, prévalence de pathologies dans le doute, etc.)*
9. Avez-vous trouvé l'intégration des résultats AI dans le rapport structuré utile? Pas du tout utile ……. Vraiment utile *[échelle 1–4]*
10. Le pré-remplissage automatique des rapports par l'AI a-t-il réduit votre temps de rédaction? Pas du tout réduit ……. Vraiment réduit *[échelle 1–4]*
11. Que manquerait-il pour vous faire gagner davantage de temps?
12. Y a-t-il des types d'examens où le compte-rendu pré-rempli était particulièrement pertinent / peu pertinent?

13. Avez-vous utilisé tout le temps la configuration V3 pour le rapport en phase 3?

- Toujours (5)
- La plupart du temps (4)
- Parfois (3)
- Rarement (2)
- Jamais (1)

14. Quelles améliorations seraient pertinentes selon vous?

15. Comment évalueriez-vous la probabilité de recommander notre produit/service/organisation à d'autres sur une échelle de 1 à 10?

Très improbable ……. Certain de recommander *[échelle 1–10]*

Annex 2. Clinical examples demonstrating the importance of AI assistance in Phase 3

Case 1

Patient: 53-year-old, post-traumatic assessment **Clinical question:** Suspected meniscal tear

Upon accessing the prefilled report, the radiologist is presented with the AI output on the PACS (Fig. 1) and with a draft report in the RIS containing all detected abnormalities prefilled in both the **Findings** and **Impression** sections (Fig. 2). This significantly reduces the time required for dictation for general radiologists and ensures consistency in reporting (Phase 3).

AI Output:

The algorithm automatically identifies and includes the following abnormalities:

- **Medial meniscal lesion** (Fig. 3a)
- Cartilage lesions of the 3 compartments
- **Bone marrow edema** affecting for instance both aspects of the femoro-patellar joint (Fig. 3b)
- **Anterior cruciate ligament (ACL) tear** (Fig. 3c)

These findings are seamlessly integrated into the radiology workflow, with all pertinent data available at first review, as shown below.

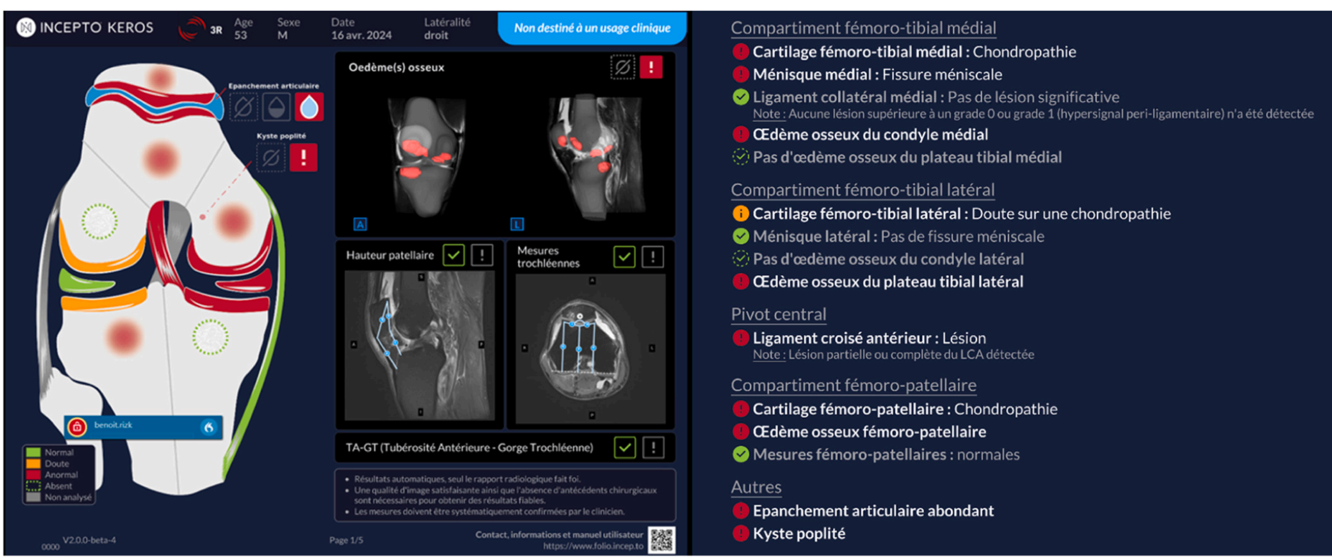


Fig. 1. AI output in the PACS

IRM du genou du 16/04/2024

Indications

[INDICATION].

<u>Technique</u> : protocole standard.

Description

<u>Compartiment interne</u> : déchirure méniscale; chondropathie; œdème osseux du condyle fémoral; pas de lésion significative du LLI.

<u>Compartiment externe</u> : pas de déchirure méniscale; doute sur chondropathie; œdème osseux du plateau tibial; pas de lésion du LLE.

<u>C. fémoro-patellaire</u> : chondropathie; œdème osseux fémoro-patellaire; rotule de configuration normale.

<u>Ligaments croisés</u> : déchirure du LCA; pas de lésion du LCP.

<u>Tendons</u> : sans altération.

<u>Parties molles</u> : kyste poplité.

<u>Synoviale</u> : épanchement articulaire abondant.

Conclusion

Déchirure du LCA. Déchirure méniscale interne. Chondropathie fémoro-tibiale interne. Œdème osseux du condyle fémoral interne. Doute sur chondropathie fémoro-tibiale externe. Œdème osseux du plateau tibial externe. Chondropathie fémoro-patellaire. Œdème osseux fémoro-patellaire. Kyste poplité. Epanchement articulaire abondant.

Fig. 2. Prefilled report in the RIS- When the radiologist opens the report, all features evaluated by the AI (normal or abnormal) are already prefilling impression and conclusion subsections of the report. Radiologist doesn't need to dictate for instance the 4 lines of the current conclusion providing time efficiency

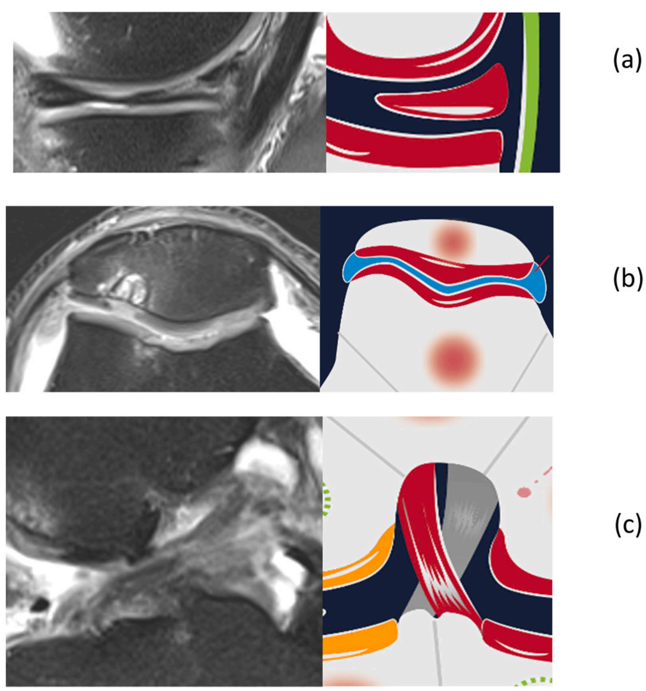


Fig. 3. (a) Medial meniscal tear detected, (b) Femoro-patellar chondral lesions with bone marrow edema on both side of the joint, (c) Complete tear of the ACL

Case 2
Patient: 64-year-old undergoing imaging for **osteoarthritis evaluation**
The AI algorithm identifies several abnormalities (Fig. 4) and directly populates them into the report (Fig. 5):

- **Popliteal cyst** (Fig. 6a)
- **Unstable meniscal tear** with a **displaced fragment** (Fig. 6b)
- **Chondral lesion** with associated **bone marrow edema** at the **trochlear groove** (Fig. 6c)

The third finding was initially missed by the general radiologist and it is depicted by the algorithm. The output is prefilled in the description and the impression of the report. Small joint effusion and popliteal cyst are also depicted by the AI.

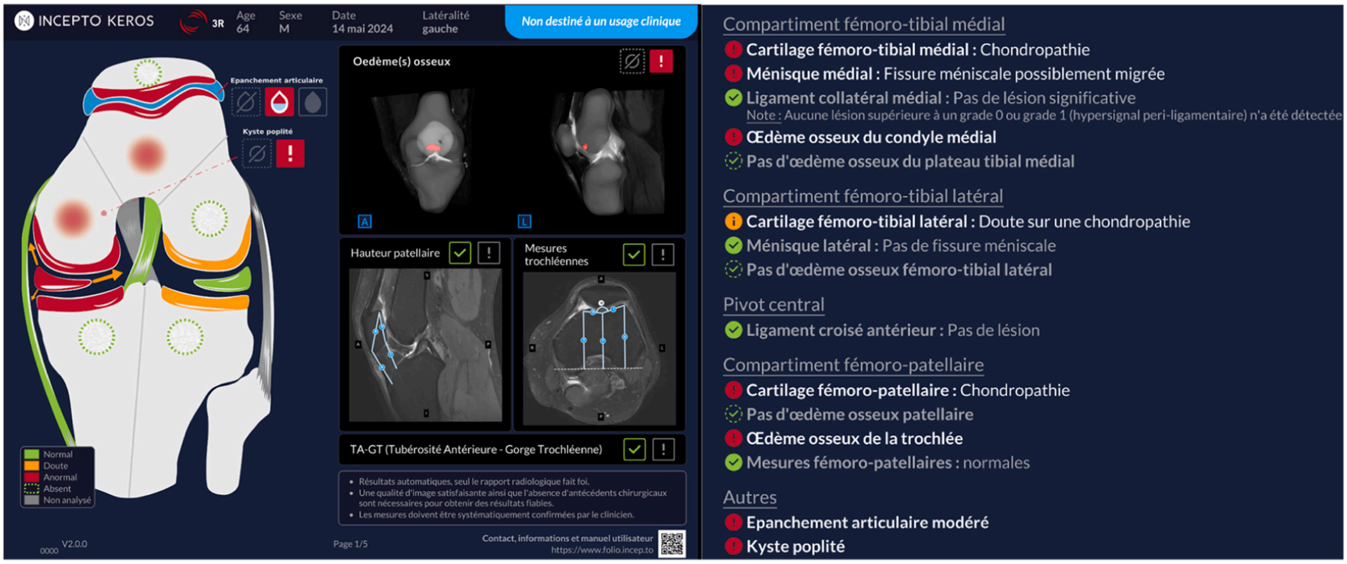
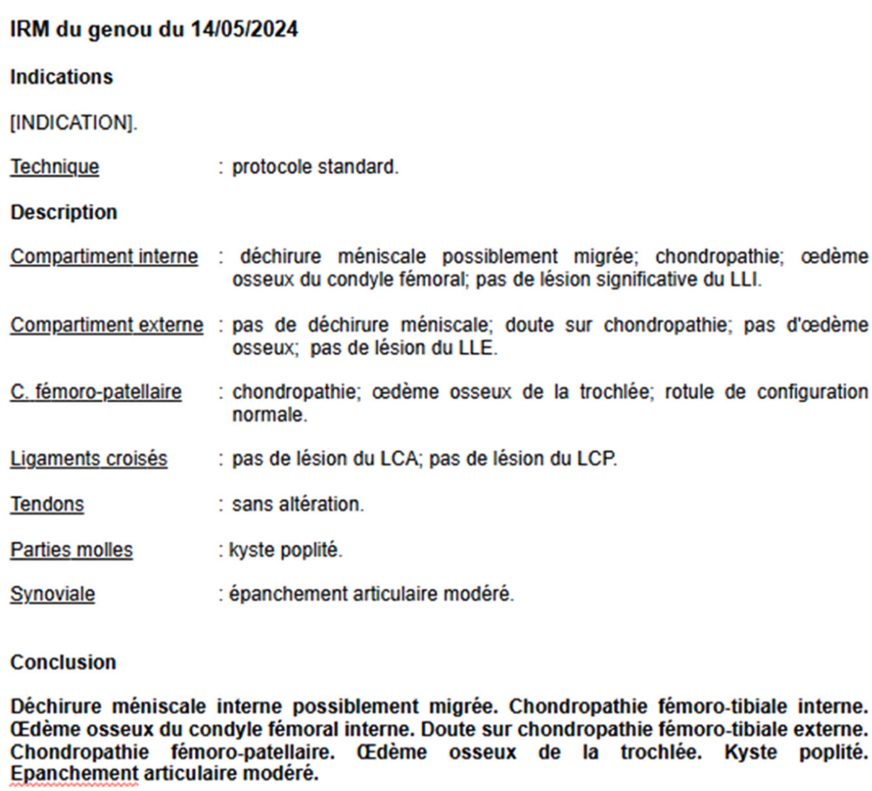


Fig. 4. AI ouput in the PACS



IRM du genou du 14/05/2024

Indications

[INDICATION].

Technique : protocole standard.

Description

Compartiment interne : déchirure méniscale possiblement migrée; chondropathie; œdème osseux du condyle fémoral; pas de lésion significative du LLI.

Compartiment externe : pas de déchirure méniscale; doute sur chondropathie; pas d'œdème osseux; pas de lésion du LLE.

C. fémoro-patellaire : chondropathie; œdème osseux de la trochlée; rotule de configuration normale.

Ligaments croisés : pas de lésion du LCA; pas de lésion du LCP.

Tendons : sans altération.

Parties molles : kyste poplité.

Synoviale : épanchement articulaire modéré.

Conclusion

Déchirure méniscale interne possiblement migrée. Chondropathie fémoro-tibiale interne. Œdème osseux du condyle fémoral interne. Doute sur chondropathie fémoro-tibiale externe. Chondropathie fémoro-patellaire. Œdème osseux de la trochlée. Kyste poplité. Epanchement articulaire modéré.

Fig. 5. Prefilled report in the RIS

The chondral lesion, missed on first interpretation, underscores the AI's potential in detecting subtle pathologies and improving diagnostic accuracy. Both the **description** and **impression** of the report are automatically updated to reflect these findings. This case further highlights the critical support role of AI in diagnostic radiology, especially in complex degenerative pathologies.

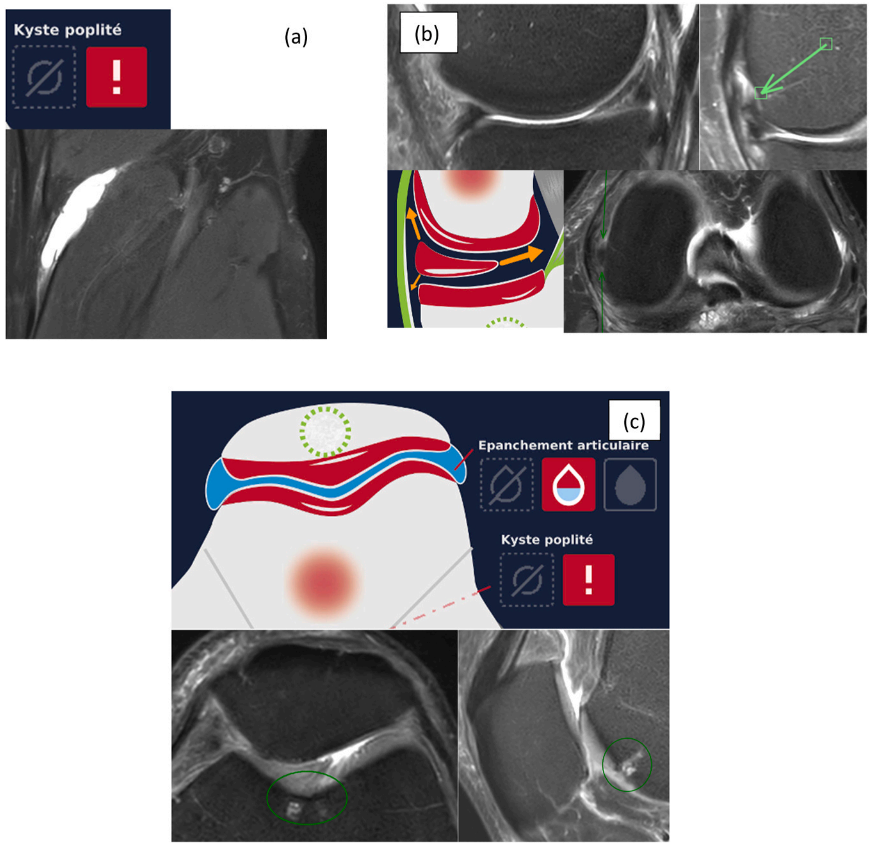


Fig. 6. (a) Popliteal cyst is identified and inserted automatically in the report, (b) Unstable meniscal tear is identified with displaced fragment (green arrows), (c) Chondral lesion with bone marrow edema of the trochlear groove (green circles)

With these findings it took the radiologist 5 min to finalize the report with the support of AI.

Data availability

Data will be made available on request.

References

[1] M. Zanardo, J.J. Visser, A. Colarieti, R. Cuocolo, M.E. Klontzas, D. Pinto Dos Santos, F. Sardanelli, Impact of AI on radiology: a EuroAIM/EuSoMII 2024 survey among members of the European Society of Radiology, Insights Imaging 15 (2024) 240, https://doi.org/10.1186/s13244-024-01801-w (European Society of Radiology (ESR)).
[2] U.J. Muehlematter, P. Daniore, K.N. Vokinger, Approval of artificial intelligence and machine learning- based medical devices in the USA and Europe (2015–20): a comparative analysis, Health Policy 3 (2021).
[3] K.G. Van Leeuwen, M. De Rooij, S. Schalekamp, B. Van Ginneken, M.J.C.M. Rutten, Clinical use of artificial intelligence products for radiology in the Netherlands between 2020 and 2022, Eur. Radiol. 34 (2023) 348–354, https://doi.org/10.1007/s00330-023-09991-5.
[4] K.G. Van Leeuwen, S. Schalekamp, M.J.C.M. Rutten, B. Van Ginneken, M. De Rooij, Artificial intelligence in radiology: 100 commercially available products and their scientific evidence, Eur. Radiol. 31 (2021) 3797–3804, https://doi.org/10.1007/s00330-021-07892-z.
[5] D.J. Blezek, L. Olson-Williams, A. Missert, P. Korfiatis, AI integration in the clinical workflow, J. Digit Imaging 34 (2021) 1435–1446, https://doi.org/10.1007/s10278-021-00525-3.
[6] T. Jorg, M.C. Halfmann, F. Stoehr, G. Arnhold, A. Theobald, P. Mildenberger, L. Müller, A novel reporting workflow for automated integration of artificial intelligence results into structured radiology reports, Insights Imaging 15 (2024) 80, https://doi.org/10.1186/s13244-024-01660-5.
[7] G. Herpe, B. Rizk, G. D'Assignies, Closing the gap: commercialized deep learning solutions for knee MRI are already transforming clinical practice, Eur. Radiol. (2025), https://doi.org/10.1007/s00330-025-11550-z.
[8] J. Behr, C. Nich, G. D'Assignies, C. Zavastin, P. Zille, G. Herpe, R. Triki, C. Grob, N. Pujol, Deep learning-assisted detection of meniscus and anterior cruciate ligament combined tears in adult knee magnetic resonance imaging: a crossover study with arthroscopy correlation, Int. Orthopaed. (SICOT) (2025), https://doi.org/10.1007/s00264-025-06531-2.
[9] B. Rizk, H. Brat, P. Zille, R. Guillin, C. Pouchy, C. Adam, R. Ardon, G. d'Assignies, Meniscal lesion detection and characterization in adult knee MRI: a deep learning model approach with external validation, Phys. Med. 83 (2021) 64–71, https://doi.org/10.1016/j.ejmp.2021.02.010.
[10] A. Tran, L. Lassalle, P. Zille, R. Guillin, E. Pluot, C. Adam, M. Charachon, H. Brat, M. Wallaert, G. d'Assignies, B. Rizk, Deep learning to detect anterior cruciate ligament tear on knee MRI: multi-continental external validation, Eur. Radiol. 32 (2022) 8394–8403, https://doi.org/10.1007/s00330-022-08923-z.
[11] M. Khalifa, M. Albadawy, AI in diagnostic imaging: revolutionising accuracy and efficiency, Comput. Methods Prog. Biomed. Update 5 (2024) 100146, https://doi.org/10.1016/j.cmpbup.2024.100146.
[12] R. Obuchowicz, J. Lasek, M. Wodziński, A. Piórkowski, M. Strzelecki, K. Nurzynska, Artificial Intelligence-empowered radiology—current status and critical review, Diagnostics 15 (2025) 282, https://doi.org/10.3390/diagnostics15030282.
[13] J. Hartung, Statistical Meta-analysis with Applications, Wiley, Hoboken, NJ, 2008.
[14] Ines Montani, Matthew Honnibal, Adriane Boyd, Sofie Van Landeghem, Henning Peters, 2023, Explosion/spaCy: v3.7.2: Fixes for APIs and Requirements. 10.5281/ZENODO.1212303.
[15] A.Q. Jiang, A. Sablayrolles, A. Mensch, C. Bamford, D.S. Chaplot, D. de las Casas, F. Bressand, G. Lengyel, G. Lample, L. Saulnier, L.R. Lavaud, M.-A. Lachaux, P. Stock, T.L. Scao, T. Lavril, T. Wang, T. Lacroix, W.E. Sayed, Mistral 7B (2023), https://doi.org/10.48550/ARXIV.2310.06825.
[16] M. Chen, Y. Wang, Q. Wang, J. Shi, H. Wang, Z. Ye, P. Xue, Y. Qiao, Impact of human and artificial intelligence collaboration on workload reduction in medical image interpretation, npj Digit. Med. 7 (2024) 349, https://doi.org/10.1038/s41746-024-01328-w.
[17] M. Cotena, A. Ayobi, C. Zuchowski, J.C. Junn, B.D. Weinberg, P.D. Chang, D. S. Chow, J.E. Soun, M. Roca-Sogorb, Y. Chaibi, S. Quenet, Enhancing radiologist efficiency with AI: a multi-reader multi-case study on aortic dissection detection and prioritization, Diagnostics 14 (2024) 2689, https://doi.org/10.3390/diagnostics14232689.
[18] B. Yacoub, A. Varga-Szemes, U.J. Schoepf, I.M. Kabakus, D. Baruah, J.R. Burt, G. J. Aquino, A.K. Sullivan, J.O. Doherty, P. Hoelzer, J. Sperl, T. Emrich, Impact of Artificial Intelligence assistance on chest CT interpretation times: a prospective

randomized study, Am. J. Roentgenol. 219 (2022) 743–751, https://doi.org/10.2214/AJR.22.27598.

[19] K. Wenderott, J. Krups, F. Zaruchas, M. Weigl, Effects of artificial intelligence implementation on efficiency in medical imaging—a systematic literature review and meta-analysis, npj Digit. Med. 7 (2024) 265, https://doi.org/10.1038/s41746-024-01248-9.

[20] H.J. Shin, K. Han, L. Ryu, E.-K. Kim, The impact of artificial intelligence on the reading times of radiologists for chest radiographs, npj Digit. Med. 6 (2023) 82, https://doi.org/10.1038/s41746-023-00829-4.

[21] M.P. Recht, M. Dewey, K. Dreyer, C. Langlotz, W. Niessen, B. Prainsack, J.J. Smith, Integrating artificial intelligence into the clinical practice of radiology: challenges and recommendations, Eur. Radiol. 30 (2020) 3576–3584, https://doi.org/10.1007/s00330-020-06672-5.

[22] F. Pesapane, M. Codari, F. Sardanelli, Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine, Eur. Radiol. Exp. 2 (2018) 35, https://doi.org/10.1186/s41747-018-0061-6.