Article published online: 2024-12-11

Review Article

# A Radiologist's Perspective of Medical Annotations for AI Programs: The Entire Journey from Its Planning to Execution, Challenges Faced

Anuradha Rao[1]

[1] Department of Clinical Science, Philips Innovation Campus, Bangalore, Karnataka, India

Indian J Radiol Imaging

**Address for correspondence** Anuradha Rao, MBBS, DMRD, DNB, Department of Clinical Science, Philips Innovation Campus, Bangalore 560045, Karnataka, India (e-mail: anu78rao@gmail.com; anuradha@philips.com).

## Abstract

Artificial intelligence (AI) in radiology and medical science is finding increasing applications with annotations being an integral part of AI development. While annotation may be perceived as passive work of labeling a certain anatomy, the radiologist plays a more important role in this task apart from marking the structures needed. Apart from annotation, more important aspect of their role is planning the anatomies/pathologies needed, type of annotations to be done, choice of the annotation tool, training the annotators, planning the duration of annotation, etc. A close interaction with the technical team is a key factor in the success of the annotations. The quality check of both the internally and externally annotated data, creating a team of good annotators, training them, and periodically reviewing the quality of data become an integral part of their work. Documentation related to the annotation work is another important area where the clinician plays an integral role to comply with the Food and Drug Administration requirements, focused on a clinically explainable and validated AI algorithms. Thus, the clinician becomes an integral part in the ideation, design, implementation/execution of annotations, and its quality control. This article summarizes the experiences gained during planning and executing the annotations for multiple annotation projects involving various imaging modalities with different pathologies.

**Keywords**

- artificial intelligence
- radiologist
- annotations
- challenges
- perspective
- data
- annotation tool
- machine learning

## Introduction

Artificial intelligence (AI) has substantial potential of applications in the field of medical imaging. Machine learning and deep learning algorithms have been developed to improve workflows in radiology or to assist the radiologist by automating tasks such as lesion detection or medical imaging quantification.[1] Annotation is an integral component of AI programs, more so in the health care domain. The annotation process aims to transfer human knowledge to the AI models by summarizing and assigning predefined labels to the digital data content. Image annotation is widely employed in medical applications, where imaging modality is annotated by an expert to improve the model's performance.[2] The annotation process is more than marking or outlining the anatomical structures or pathologies. It is a process that is

**DOI** https://doi.org/10.1055/s-0044-1800860.
**ISSN** 0971-3026.



Thieme Medical and Scientific Publishers Pvt. Ltd., A-12, 2nd Floor, Sector 2, Noida-201301 UP, India

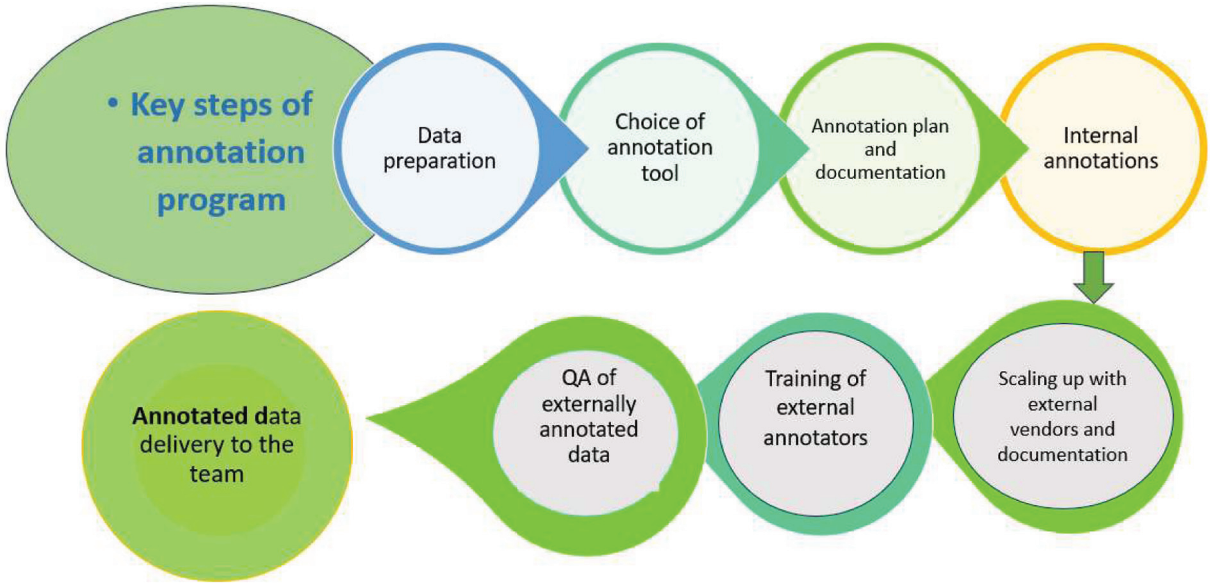


**Fig. 1** Important steps in the annotation program. QA, quality assurance.

more elaborate and challenging as a clinician's role is involved in the entire process of data creation for annotation, annotation planning, annotation execution, training the annotators, scaling up of the annotation with an external vendor, quality control (QC), etc. In this article, the experience gained from annotating for multiple AI projects has been summarized. The end-to-end annotation process from its inception, implementation to handover of annotations of AI projects has been described in this article, which highlights the important role of the physician/radiologist in the design of the annotations to its final delivery to the technical team for development of AI solutions.

The entire annotation program can be grossly categorized into multiple steps as below (►**Fig. 1**):

1. **Data preparation:** While data preparation could be part of the technical team's work, involvement of the physician would be immensely helpful in controlling the quality of data annotations. Looking for variations in the data are of prime importance as variations are the key for the training of AI algorithms.
   Based on the nature of data used in the project, we can divide the data into two categories (►**Fig. 2**):
   a. **Data obtained from planned clinical study:** In this method of acquisition of data, the clinician would have more control on the data. The study can be planned as per the requirements of the study and needs of the output planned and can have data variations planned. Introduction of different variables is an important aspect of data collection as every AI algorithm needs

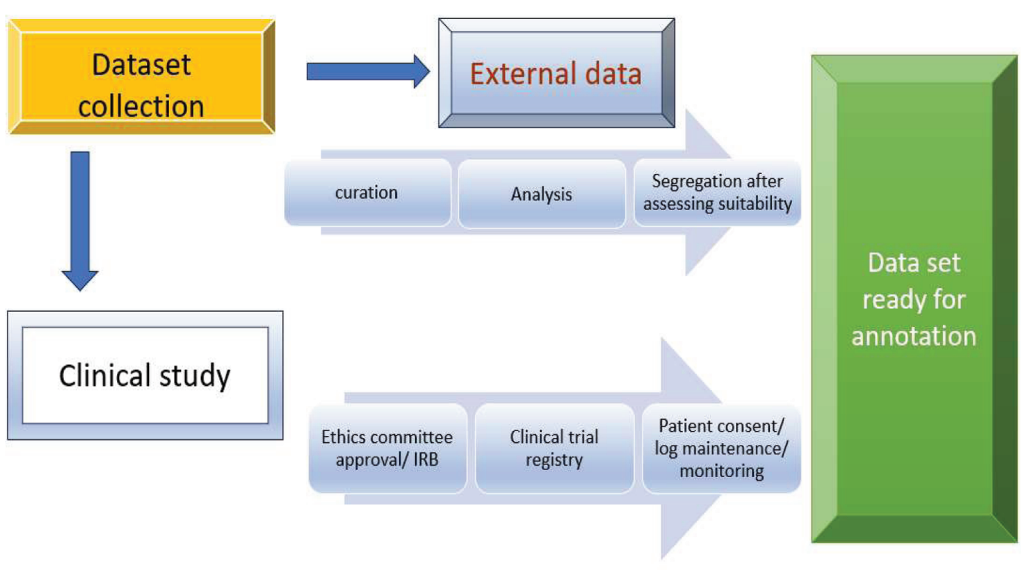


**Fig. 2** A schematic depiction of the dataset creation/collection. IRB, institutional review board.

 

a combination of variables of patients, machines, settings, and postprocessing algorithms, which can be extremely difficult to obtain.[3] This will need immense planning of the type of data needed, method of acquisition, protocol for the data acquisition, compliance with the protocol, ethical considerations, etc., which are beyond the scope of this article, but are sure enough to involve the clinician's expertise in all the mentioned aspects of the study. Considering the inclusion of variations in the planning phase of the study is important to obtain variations in the datasets. For example, planning a data collection study for a lung cancer screening program by chest radiographs for an automated lung nodule detection would involve data to have nodules from benign to malignant nature, along with inflammatory and infective mimics of the nodules. These are important as negative class training labels for the algorithm.

The clinical curation of the acquired data from the clinical sites would be a definite prerequisite for preparing the data for annotation, which again rests on the radiologist and his or her team to finely filter the unwanted data and provide a clean, needed, protocol compatible data that will be de-identified and set in pipeline for annotations. Ideally this exploratory data analysis should be conducted as a joint collaborative effort from the clinicians, statisticians, data scientists, and developers.[4] The documentation of data curation is a prerequisite for Food and Drug Administration (FDA) submission purposes. It will be a good idea to craft this document in detail in alignment with the FDA requirements.

b. **Data obtained from any other source:** Any method of data acquisition other than the planned clinical study will have more curation to be done as the data may not directly match with the exact requirements of the project in question. Cataloging of the data becomes an important work for the clinical team. Analyzing the data that can be compatible and usable with what the specific project needs is a big task. For example, the obstetric data that have been collected solely for gestational age may not be directly applicable to other anatomical structures such as the placenta and amniotic fluid. In this scenario, the data need to be analyzed to determine if the whole uterus has been assessed for placenta localization and for amniotic fluid evaluation. Similarly, all the chest radiographs used for health checkup purposes may be used for screening of lung nodules; however, radiographs collected from an oncology center may induce a bias. Hence, analyzing the data and their usability is of prime importance before planning any annotations. The prerequisites for clean data and the methods of data curation need to be documented as a standard operating procedure (SOP) for each data to be loaded for annotation.

The de-identification process is a must before any data are loaded to the annotation tool. It is a very important step in compliance with regulations to ensure patient privacy (strict anonymization/using encrypted platforms, etc.). This is usually handled by a separate team, which ensures patient privacy. Data security is another area of regulatory significance that can be made compliant by ensuring access to authorized persons only, ensuring secure data transfer, storage, etc.

The training and validation datasets need to be set aside from the available data. The neural networks will be "trained" using training datasets from which it "learns." These are usually hand-labeled image datasets. Once a network has been trained, it will be tested on a different set of data (validation datasets), designed to evaluate the model on new data.5

2. **Choice of annotation tool:** It is imperative to have a thorough investigation of the available annotation tools that will cater to the needs of the annotators to fast track their annotation workflow. From requirement gathering to scouting for the appropriate tools, identifying whitespace, and, finally, sorting to meet the defined needs such as basic annotation functionalities, collaborative work (web-based, traceability), data formats (in, out, and interims), and finally affordability. The choices are ample. Choosing a right annotation tool would reduce the amount of work and time needed for annotation.[2] There are many free open-source annotation tools available, which can be made use of for cost reduction. Paid third-party annotation tools are available, which can be up and running quickly. They have complete data auditability and compliance and may provide some customization for the project concerned. Validated annotation tools may be a prerequisite for FDA submission. Multiple factors must be considered while choosing an annotation tool. Understanding of the annotation needs, the cost involved, speed of annotation provided by the tool, ease of annotation, storage facilities, accessibility to the annotators depending on their location (internal versus external), etc., will determine the annotation tool that will be used. The annotations can be performed in a standard shape, which can be circle, point, freehand-drawn mask, ellipse, polygon, polyline, etc. There can be temporal segments, which are determined by beginning and end timestamps.[2] The planned annotation needs to be tried out on these different tools to look for ease of annotation and speed of annotation. Some of the key elements to speed up the annotation like interpolation tool needs to be checked for accuracy. Some of the annotation tools offer identification of the annotator's name, which can be very helpful in a QA process. There are different image adjustments, measurements, rendering pre-sets, and color maps that make the labeling process easier and more accurate. These need to be understood properly before the tool is finalized for annotation based on the requirements for the concerned project. The tool should have options for data encryption and anonymization, and meet legal requirements. Some of the institutes recommend validated annotation tools to comply with the future needs of the FDA.

 

3. **Annotation plan and documentation:** Once the data are collected and reviewed and curated, an annotation plan that has good reproducibility and can be completed in the planned timeline of the project must be charted out.[6] At the outset, an overview of the outcome of the project would be the crucial factor in deciding the anatomy or the pathologies to be identified in the given clinical dataset. For example, if the outcome is just gestational age for an obstetric AI project, all the anatomies that are related to biometry need to be planned for annotations. Similarly, if the end result of the AI-generated output would be intracranial hemorrhage on a CT scan, all the anatomical structures related to the possible location of the hemorrhage and mimics of hemorrhages like calcifications should be planned for annotations. Listing of the needed anatomical structures is the first step in the annotation process. This will give us the least number of anatomical structures needed for annotation. While the increased number of anatomical structures would be beneficial for technical development, the cost of annotation would also be increased with the increased time for annotation. The radiologist needs to optimally select the most important anatomies relevant for the technical feature to be developed. This will reduce the cost of annotation and the time needed for annotation.

   Another important component for annotation would be negative classes of anatomy or pathology, which are very important for AI training. For example, in a planned biometry of abdominal circumference, annotating kidneys would be a negative class. Documenting the plan and discussing with the technical team would be an ideal beginning. All these plannings require an understanding of the importance of the anatomical structure in question and also sometimes its pathological significance, which requires a clinician's knowledge and experience. The perspective of annotation of the technical team is just as important as that of the clinician as they can give valuable information on annotation planning based on technical insights.

   The annotation plan depends on the type of imaging modality being annotated. For example, most of the computed tomography (CT) scans have fixed imaging planes without significant movement artifacts unlike the fetal ultrasound. A project intended to detect chronic infarcts on magnetic resonance (MR) may be simpler than a project that involves fetal ultrasound. While the former has stable images acquired from MR of different sequences, the latter has multiple ultrasound videos that involve maternal and fetal anatomical parts with movements related to the probe and the fetus. With respect to various pathologies that need to be annotated, the anatomy that needs to be annotated depends on the project content and goals. Some examples include tagging pneumonia on a chest X-ray or annotations of various interstitial lung diseases on a high-resolution CT scan.

   The choice of the annotation based on detection or classification needs to be decided by consensus with the technical personnel. The annotation process itself may be a very time-consuming task, especially for specific annotations like the segmentation task, which is very labor intensive and can be expensive.[7] While segmentation (►**Fig. 3**) may be the best option for many anatomical structures without definite geometric shapes, it can be heavy to process too on certain smaller devices like portable machines. Lighter options like bounding boxes (►**Fig. 4**) and tagging (►**Fig. 5**) may be preferred over segmentation in some projects handling low processors. The segmentation can be used to create two-dimensional (2D) and three-dimensional (3D) models, which can give valuable information (►**Fig. 6**). The annotations can be done for various modalities like X-ray, ultrasound, CT, and magnetic resonance imaging (MRI). The figures and images provided in this article are from annotations/segmentations done from various annotation projects (►**Fig. 7**).

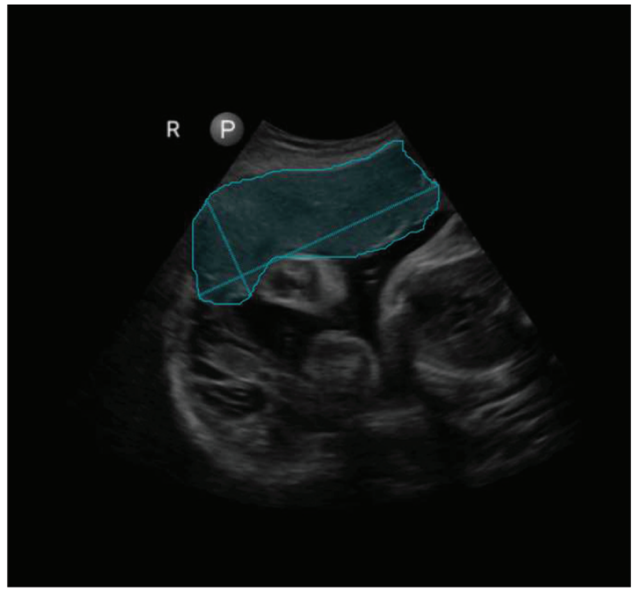


**Fig. 3** Segmentation of placenta.

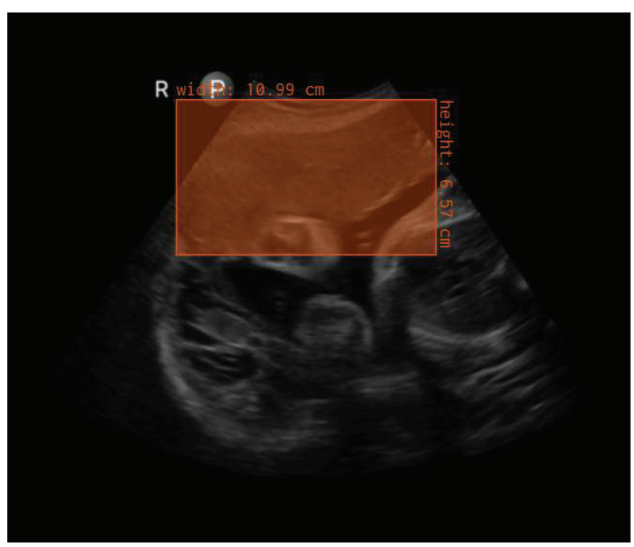


**Fig. 4** Bounding box annotation of placenta.

 

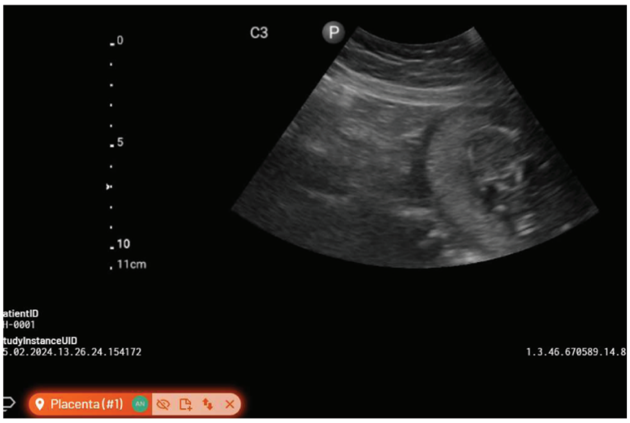


**Fig. 5** Tagging the image as placenta.

Once the annotation planning is finalized, a small subset of chosen data could be used for a pilot annotation program and tested for its usability.

The most important aspect of the pilot annotation is the feedback from the technical team on any modification that needs to be done from the annotation standpoint. This is a crucial point where fine-tuning of the annotations begins. It should always be kept in mind that these are research projects; trial and error of different methods and anatomies of annotation would be required at any point of the annotation program. Both the annotators and the technical team should have a mutual understanding of this fact at all times. Mental fatigue or frustration when a revision of the annotation is suggested can be tackled by preparedness for such a situation. The fact that a final decision can be made only by multiple or at least a few attempts should be mentally accepted by both the annotators and the technical team. While the clinicians understand that human science is not black and white and have shades of gray, the same may not be perceived by the engineering team. It is important to make the technical team understand that the clinical science is very variable,

**Fig. 7** Segmentation of prostate on magnetic resonance imaging.

and every human body is unique and the same may reflect in the AI of medicine as well.

Proper documentation of this annotation plan is an important step toward progress in the annotation process. Apart from serving as a document needed for submission for FDA approval, it serves as a reference tool for future.[8] Different versions can be documented as and when new modifications or additions of new labels are implemented. This is very crucial as the participants in the annotation, technical team involved, and human resources involved in the project may change over the course of the project.

4. **Internal annotations:** Involvement of the appropriate clinical specialists is a basic requirement in a medical AI project. Involvement of external medical experts can be taken on a need basis for scaling up of annotations. However, a few of the internal clinical team members are needed for larger AI projects who can plan the annotation, guide through the annotation process, and be critical QA experts of the annotated data. As mentioned earlier, a pilot internal annotation needs to be done to get feedback on the usability of the planned annotation. Depending on the number of internal clinical team members, the number of data to be annotated can be planned

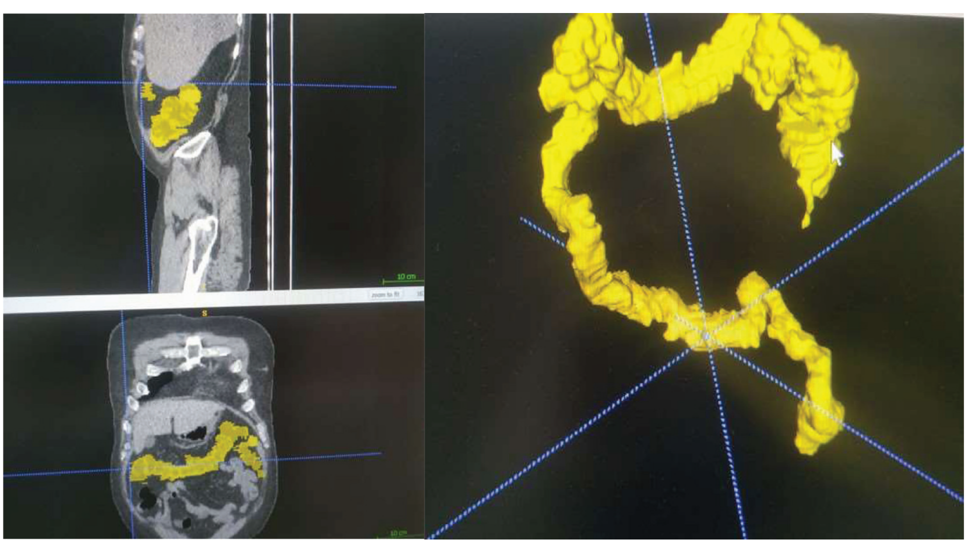

**Fig. 6** Segmentation of the large bowel done on the coronal, sagittal, and axial (not shown) image yielding a 3D depiction of the colon.

 

after internal training. There should be a clear consensus on the type and nature of annotations between the annotators. Interobserver variations do exist despite the best efforts as per the clinical annotation plan. This needs to be considered after screening for the internal annotator's initial annotations. Having different annotators annotate the same anatomies would be one method of introducing interobserver variations in the training set. These are observed more often in some labels like orientation labels, and when annotation is performed in the suboptimal images like shadowed areas or in images with artifacts, etc. For example, a spine sagittal image may be interpreted as an oblique image by another annotator when the image is not exactly in the true sagittal plane. Similarly, motion-blurred CT images may induce interobserver variations while performing measurements or segmentations of a lung nodule.

The advantages of internal annotations include the following:

- Flexibility of the annotators to adapt to any changes in the annotation plan.
- Corrections can be easier.
- Project plan visibility in the annotator team can be of immense help for better annotation.
- Instant feedback can be given to the technical team.
- Relatively less expensive.
- Extensive training programs, contracts, or documentations will not be necessary. Basic annotation SOP/documentations needed as per the project/FDA requirements will be sufficient.

The disadvantages of internal annotations are the following:

- A large project with substantial data annotation requirement may make feasibility of internal annotation difficult. Scaling up to the external vendors is the only solution to complete large projects on time.
- Limited number of internal annotators may not be able to work on project deadlines.
- Sudden attrition can result in delays of annotation, which may affect the project immensely.

5. **Scaling up with external vendor and SOP/related documentation**: Scaling up of annotation to the external vendors may be the only option when the annotation is of a large scale and cannot be completed internally. The choice of vendors available in the market needs to be carefully assessed based on the medical annotation requirements. Vendors offer different strata of annotators ranging from highly skilled subspecialists in various clinical fields to general annotators. It is always good to understand that good annotations are the basis for good AI training. It is needless to say that corrections of the annotations are more cumbersome and time-consuming than annotations done well at the first instance. Hence, it is always wise to choose the relevant specialist to annotate the anatomy considering that medical/radiology annotations are difficult for general annotators and nonspecialized medical physicians. This will improve the overall quality of annotations and will result in less errors. Also, considering the FDA requirements, choosing the right qualification of annotators will lead to little or no hassle during FDA submissions later.
The number of annotations required needs to be carefully understood and planned before hiring the external annotators. The annotation needs to be planned carefully with respect to the available annotator team, distributed properly among specialists and nonspecialists depending on the needs of the project.
The time taken for annotations can be estimated so that the cost, turnaround time, time needed for completion, etc., can be planned well ahead. The final sign off of all the annotations should be by a qualified (expert) physician as per the FDA standards.
It is always better to inform the vendor that the annotations are subjected to changes as and when the algorithm requirements change, or improvements happen.
The advantages of outsourcing the annotations include the following:
   - Steady flow of annotations is possible when there large-scale data are available for annotation, which cannot be met by small internal annotator team.
   - Turnaround time of annotation could be much higher due to a large team working on the annotations round the clock.
   - Scaling up of annotations is easier when the annotations are outsourced as there is no need to recruit annotation specialists on a need basis. The annotation service providers have a talent pool that can be utilized when needed.

   The disadvantages of outsourcing the annotations are the following:
   - Training the various levels of skilled and unskilled annotators is in itself is a huge process.
   - Since the annotators are external to the project, the level of understanding of the needs of annotation will be much less. Hence, the expected quality may not match the internal annotators.
   - QC of the outsourced annotations is a big challenge.
   - Separate time needs to be set aside for writing training modules, training the annotators, and QC documentation.
   - Attrition within the outsourced companies would result in intermittent introduction of new annotators who need to be retrained.
   - Overall, it is relatively expensive compared with internal annotations.
6. **Training of annotators:** Training of annotators is a big task (►**Fig. 8**). Training the specialists and nonspecialists may be a different experience altogether. It is always good to train the specialist radiologist of the external vendors who will in turn train the general annotators whenever there is a need to hire the general annotators. A detailed preparation of the training document is a must before starting the training. Illustrations are a very important part of the document. Each anatomical structure to be annotated needs a description, which includes the type of

 

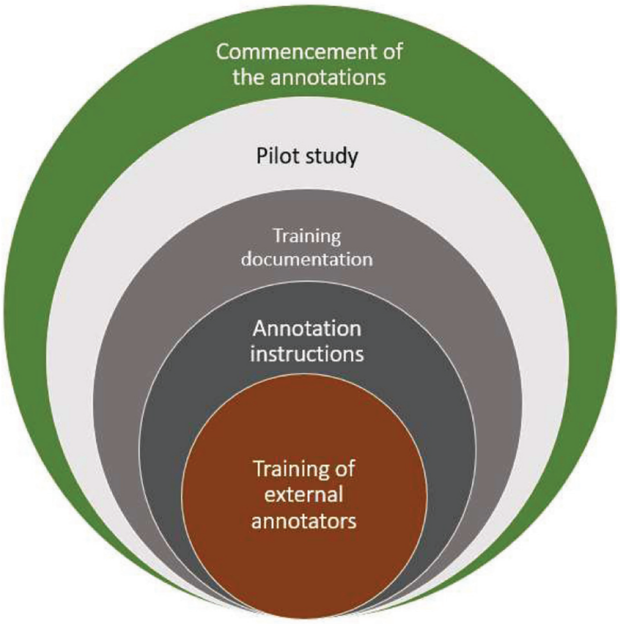


**Fig. 8** A schematic depiction of the training process of annotators and initiating the annotation program.

annotations needed with corresponding illustrations. This can be achieved by performing a few internal annotations and trying out within the internal team of annotators. This would give an insight into the possible errors that can happen during the annotation process. The training documents can be handed over to be read and comprehended before a live session of discussion with the annotators. Once the training is completed and the annotators have understood the process well, a documentation of this training needs to be done with the annotators acknowledging the same. A pilot study can be conducted where a few tasks are assigned to the future annotators. This will give confidence for the annotators and an understanding of the quality of the annotations delivered. Ongoing training may be needed during the initial weeks to improve the quality of the annotations.

7. **QA of the externally annotated data:** This is an extremely important step before the data are handed over to the technical team for development. Based on the complexity of the problem space, the annotations are the critical building blocks of the relevant algorithms. Thus, mistakes in them pose a critical challenge in the final throughput. Manual quality checks are a mandatory practice but scaling them up and making them foolproof is certainly a challenge. Tools like FiftyOne or V7 (not limited to) provide automated assistance in identifying these common mistakes in annotations, making the QA process less tedious.
   It is wise to keep the QA plan (►**Fig. 9**) handy even before starting the external annotations. An internal QA mechanism can be encouraged to be set up within external vendors to deliver quality annotations. Guidelines on the expectations related to the quality of annotations needs to be discussed with the technical team. Since there are no set standards available for acceptance of errors in the literature, it is extremely difficult to correctly predict the error percentage. Expected error percentage can be decided after consulting with the technical team based on the required results. For example, the acceptable error of a bounding box size for a larger anatomical structure may be higher than the acceptable error for a smaller anatomical structure where the precision of the size of the bounding box is critical. Similarly, segmentation requires

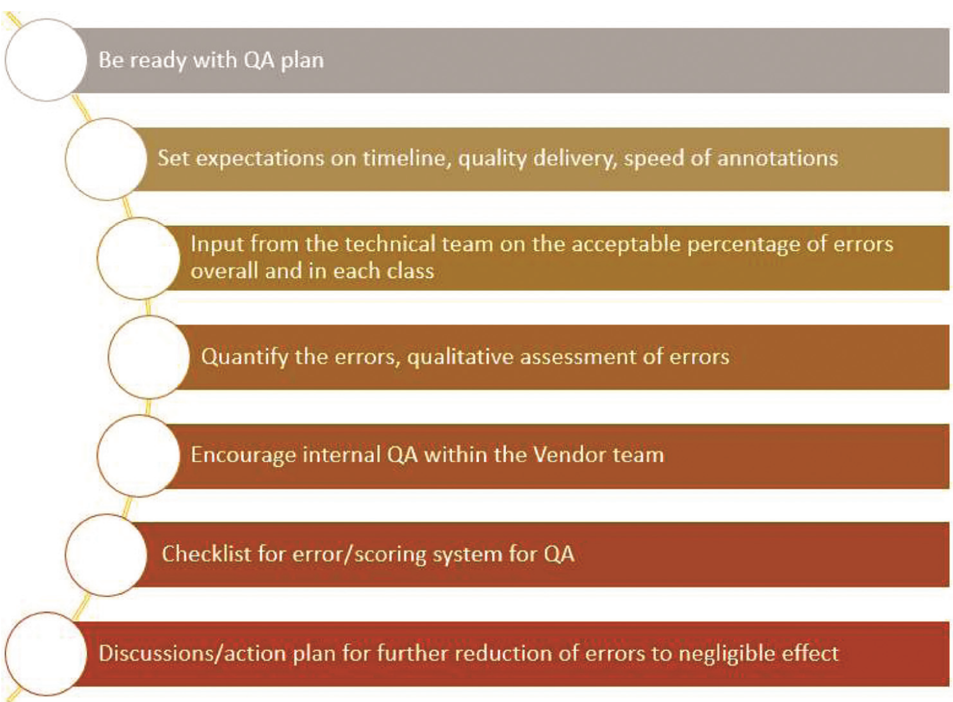


**Fig. 9** The flowchart of the quality control of the external annotations.

 

more precise annotations than the bounding boxes. The measurement-related annotations need extreme precision. Hence, based on the anatomy measured, the error margins of measurements can be decided. For example, measurement of a tumor cannot be compromised as its size may be deterministic of decision-making between surgery and conservative measures. Similarly, biometry in obstetric scans have very less error margins as they determine the growth of the fetus. In contrast, a large anatomical structure like the placenta may have a slightly higher error margin, which can be accepted as the size of the bounding box is of less importance than the location of the bounding box in an image. Similarly, a large infarct or hemorrhage in the cerebral hemisphere on a CT scan or MRI of the brain can have a relatively higher error margin of the size of bounding box compared with a small infarct at the pons or the medulla.

A checklist needs to be prepared regarding the size-related errors, anatomy-related errors, missing anatomy, measurement-related errors, etc. Scores can be assigned based on the checklist. Based on the scores, the decision of accepting or rejecting the annotation can be made.

8. **Annotated data delivery to the team:** The annotated data will be finally delivered to the technical team for algorithm development. The data delivery could happen in a platform that has direct access to the developers who can use the data directly from the platform. Alternatively, a few annotation platforms need the data to be downloaded before they can be used by the developers. The data need to be segregated according to the needs of the project.[9]

## Future Trends

### Annotation Automation

The primary objective of the transition from manual annotations (via experts or crowdsourced) to semi-automated (active or transfer learning) annotations is to focus on more complex cases, improving algorithm throughput with minimal human intervention. With more advancement in AI and related areas of annotation prediction, human annotators are assisted by suggesting regions of interest and predicting the appropriate labels based on learned data. In a curated way, this ecosystem will evolve to support the annotator community to bring more quality throughput of this raw material, boosting algorithm development to its peak.[10–12]

### AI in Quality Assurance

AI may be used to perform QA of the specialist-annotated images too. Once a sufficient amount of data has been annotated, the annotators are trained, and the AI model works well, the predictions generated by the AI model itself may be used to analyze the segmentations or detections annotated by the radiologists. Any discrepancies between the two may be referred for adjudication by an independent radiologist. This serves as AI-guided QA for the annotations/segmentations done by the specialists and saves time, resource, cost, and effort of QA for annotations.

## Conclusion

The annotation process is an integral part of AI development. The radiologist or the clinician involved in the process plays an important role in the entire process from its inception to final delivery of the annotations. Planning of the needed annotations and choosing the correct annotation platform based on the needs of the project are crucial for the success of the annotations. While internal annotations can have a strong hold on the quality of the annotations, external annotations could speed up or scale up the annotations significantly at the same time expecting stringent overseeing of quality. A good coordinated approach between the clinical and technical team is key to an effective annotation program.

## Highlights

Annotation is crucial for development of an AI solution. The radiologist plays a crucial role in the annotation process from its inception to final delivery.

**Funding**
The author currently works for a project funded by Gates foundation under grant ID :INV-058274.

**Conflict of Interest**
None declared.

### References

1 Willemink MJ, Koszek WA, Hardell C, et al. Preparing medical imaging data for machine learning. Radiology 2020;295(01):4–15
2 Aljabri M, AlAmir M, AlGhamdi M, Abdel-Mottaleb M, Collado-Mesa F. Towards a better understanding of annotation tools for medical imaging: a survey. Multimedia Tools Appl 2022;81(18):25877–25911
3 Larson DB. Openness and transparency in the evaluation of bias in artificial intelligence. Radiology 2023;306(02):e222263
4 Rouzrokh P, Khosravi B, Faghani S, et al. Mitigating bias in radiology machine learning: 1. Data handling. Radiol Artif Intell 2022;4(05):e210290
5 European Society of Radiology (ESR) What the radiologist should know about artificial intelligence: an ESR white paper. Insights Imaging 2019;10(01):44
6 Mongan J, Halabi SS. On the centrality of data: data resources in radiologic artificial intelligence. Radiol Artif Intell 2023;5(05):e230231
7 Galbusera F, Cina A. Image annotation and curation in radiology: an overview for machine learning practitioners. Eur Radiol Exp 2024;8(01):1
8 FDA. Clinical Data for Premarket Submissions. Accessed November 19, 2024 at: https://www.fda.gov/medical-devices/premarket-submissions-selecting-and-preparing-correct-submission/clinical-data-premarket-submissions
9 Yang Y, Li R, Xiang Y, et al. Expert recommendation on collection, storage, annotation, and management of data related to medical artificial intelligence. Intell Med 2023;3(02):144–149
10 Wang S, Li C, Wang R, et al. Annotation-efficient deep learning for automatic medical image segmentation. Nat Commun 2021;12(01):5915
11 Ma J, He Y, Li F, Han L, You C, Wang B. Segment anything in medical images. Nat Commun 2024;15(01):654
12 Renard F, Guedria S, Palma N, Vuillerme N. Variability and reproducibility in deep learning for medical image segmentation. Sci Rep 2020;10(01):13724

 